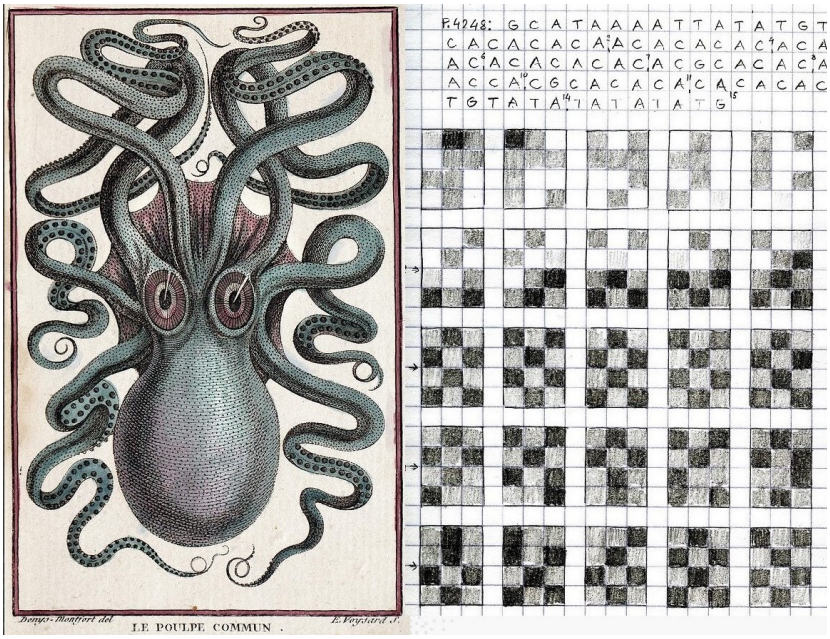
The DNA States of Micro-entropy

From Chaos to Order and Back – Part II



Ludwig Boltzmann defined entropy as a measure of the number of possible microscopic states (microstates) of a system in thermodynamic equilibrium, consistent with its macroscopic thermodynamic properties, which constitute the macrostate of the system. A useful illustration is the example of a sample of gas contained in a container. The easily measurable parameters volume, pressure, and temperature of the gas describe its macroscopic condition (state). At a microscopic level, the gas consists of a vast number of freely moving atoms or molecules, which randomly collide with one another and with the walls of the container. The collisions with the walls produce the macroscopic pressure of the gas, which illustrates the connection between microscopic and macroscopic phenomena. (Wikipedia)

In one of my earlier posts I mentioned how in my high school Physics class entropy was explained in these thermodynamic terms as a room divided in two equal parts; in one part the air was hot while in another it was cold.

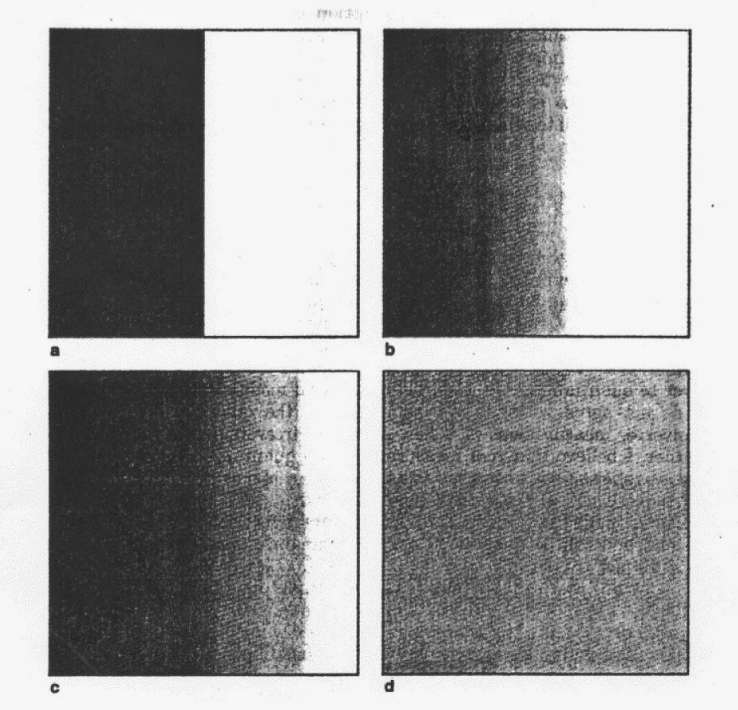


Fig.01 Boltzmann's thermodynamic model

Using a thermometer in the room we will know exactly in which of its two parts it is.(Fig.01) When the divide is removed the air in the room will gradually begin to mix and after some time the temperature in the entire room will become the same – warm. Now it would not be possible to know where the thermometer is placed.

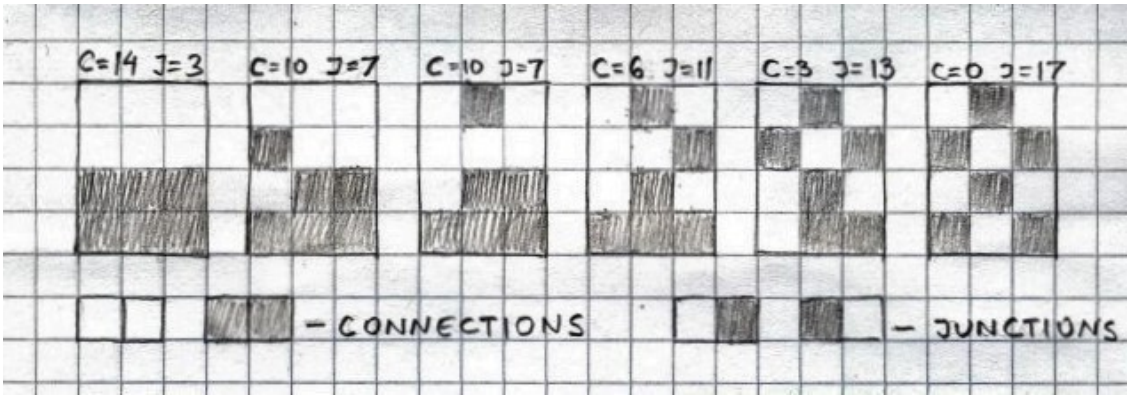


Fig.02 Transition from high to low organization

Many years later while working on a visual algorithm for representing RNA and DNA strands as 2D images, I began noticing that some in this way generated states/images have properties that could correspond to states of high and low

organization(entropy)(Fig.02). Analogous to the thermodynamic concept of entropy, most of these identified cases were binary states based on two values: dark and light.*)

However, unlike the microstates in the Boltzmann's model which are based on vast number of freely moving atoms or molecules, which randomly collide with one another and with the walls of the room, we have here DNA sequences with a small number of bases precisely defined by their positions and values attached to them. Here DNA bases as smallest units are playing roles similar to the atoms and molecules in the gas.

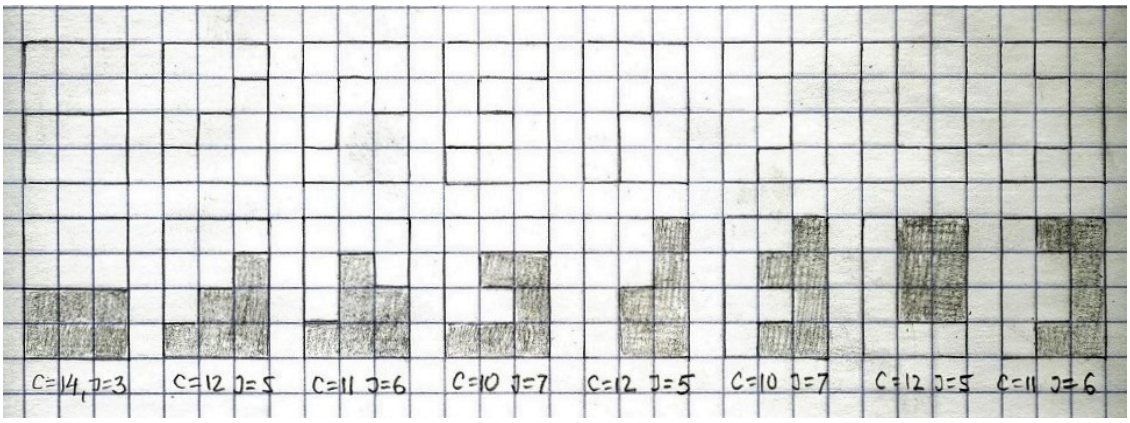


Fig.03 Eight states of high organization(6T+6A)

The notion of entropy on this micro-level is defined within a 3x4 matrix as a structure of positions and structure of values representing five RNA/DNA bases(T=white, G=light, A=gray, C=dark,U=black). For each state, representing RNA or DNA sequence, the maximum number of values is four. Since each of 12 positions in a 3x4 matrix can be occupied by a single value(base), those values would get into the spatial(2D)neighborhood relationships. If two neighboring positions have the same value,this relationship will be called „connection" (C), while the neighborhoods of different values will be „junctions" (J). In special cases of binary sequences(ATATTTATAA...) with equal numbers of both values(T=6, A=6), the states of higher organization are defined by maximum number of connections(C=max) and minimal junctions(J=min), while the states of high entropy are reverse, they are with highest number of junctions(J=max)and lowest number of connections (C=min). There are 8 different ways to divide a 3x4 matrix in two equal size parts(Fig.3). Perhaps it is worth noticing that in these

cases high organization usually appears as a single state while entropy could have several successive states. Examples of DNA sequences showing transition from the state of high organization to the state of low organization(Fig.04)

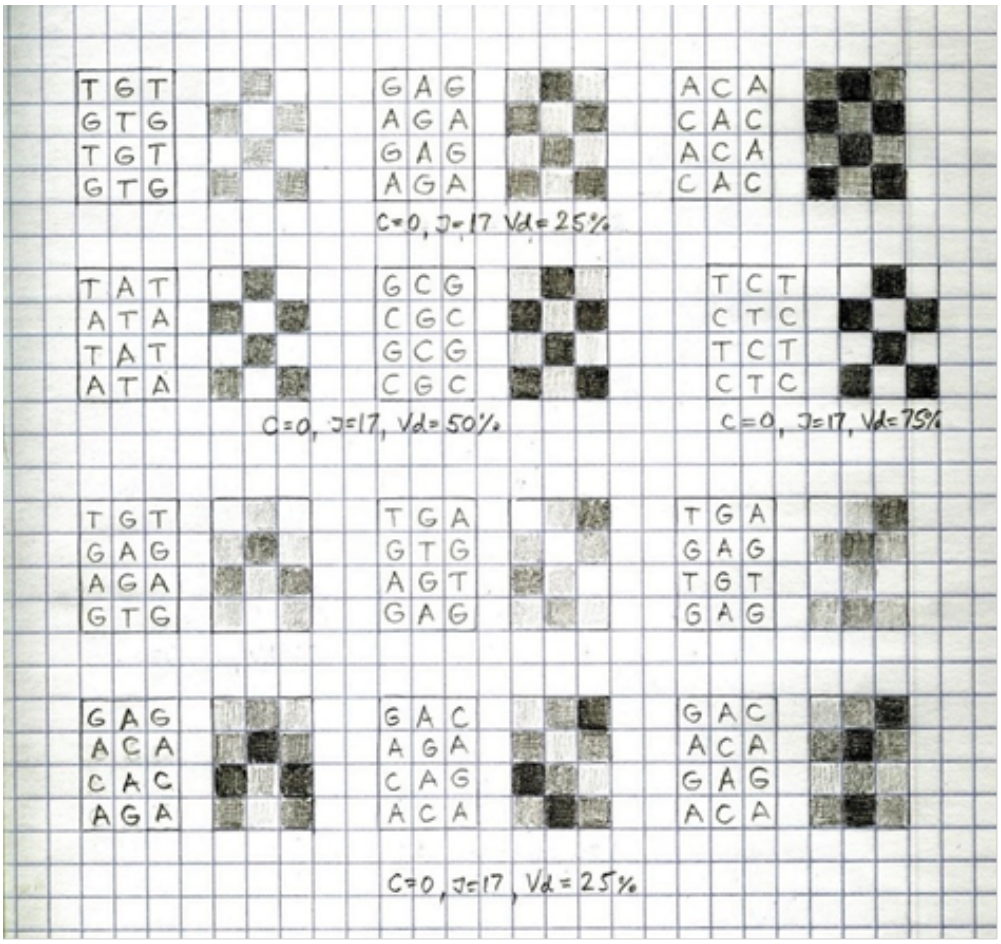


Fig.4 Eight states of high organization(6T+6A)

However, while in the case of connections(T-T, A-A) there is no value difference between the neighboring elements(Vd=0), in case of junctions it could vary (T-G=25%, T-A=50%, or T-C=75%). While in case of macro-entropy, with only two different values defined on macro level(temperature), on the micro-level there are more nuances. For example, how are different states of entropy in cases of binary sequences with different value relationships(T-G, T-A, T-C, G-A, G-C, A-C), or how to define entropy with unequal number of bases/values(T=8, A=4, or G=2, C=10)? Then, what about the states generated by sequences with 3 bases (TGA, TGC, TAC, GAC) or by all four(TGCA)?

There are DNA sequences when, converting into images, produce states in which no same values are neighbors, in other words, the number of connections C=0. For example, with three different binary sequences TGTGTGTG..., TATATATA... and TCTCTC... we will get three similar states of entropy but each with its own value differences for junctions between neighboring elements(Vd1=25%, Vd2=50%, Vd3=75%). By definition these are all states of entropy but they are not the same. Perhaps this could be compared with macro-entropy cases when the temperature difference between two rooms vary, could be lower or higher. While this might make sense in the cases of binary sequences, the comparison becomes more difficult with sequences consisting of three values/bases: TGATGATGA..., or TACTACTAC...

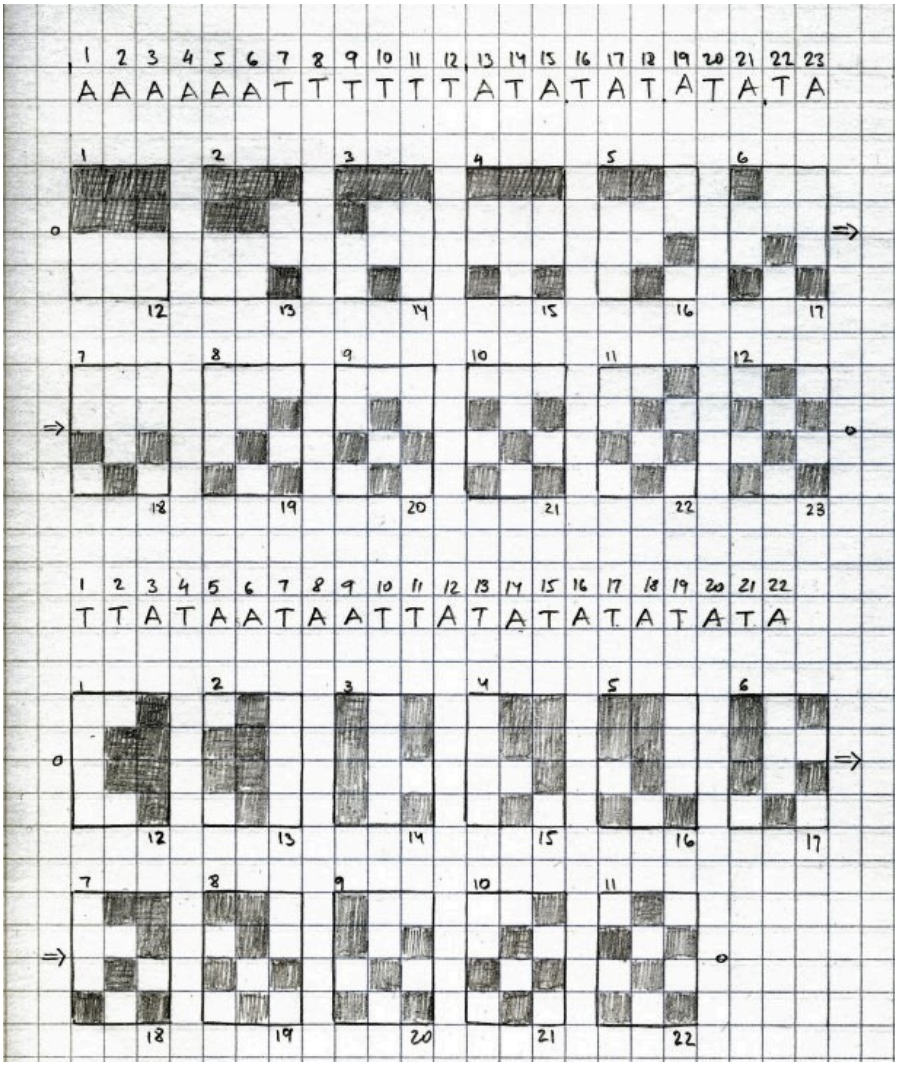


Fig.05 Transitions from high to low organization states

Recently, while working on the *Octopus sinensis* genome, which appears to be very interesting, I noticed numerous cases corresponding to the state of entropy, with various distributions of basis all having only junctions (C=0), with the same or with different value differences. For example, among its binary sequences one could find not one but three different states of entropy all based on junctions of 25% value difference(Vd=25%): 6T+6G, 6G+6A, 6A+6C.Then, there are two similar states but with Vd=50%: 6T+6A and 6G+6C, while there is only one state of entropy with Vd=75%: 6T+6C.(Fig.05)

In the *Octopus sinensis* genome there are numerous sequences with three bases that, when converted into 2D image, have only junctions, a property related to the state of entropy. In these cases entropy is not defined as a mixture of two but of three different values.(Fig.06) And, unlike the thermodynamic entropy, many DNA sequences shows reversibility, moving from higher to lower degree of organization and then back to the higher, as it could be noticed in the octopus DNA (Appendix 2).

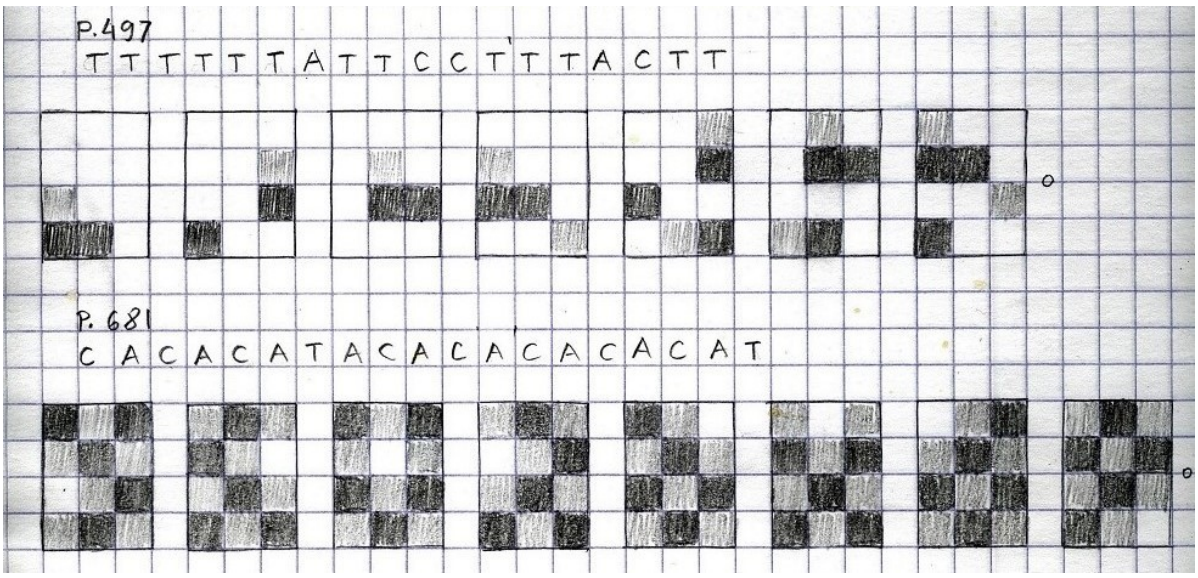


Fig.06 Octopus sinensis - DNA sequences consisting of three different bases.

Images generated by this particular algorithm are not just a better way to represent RNA and DNA and study their formal properties but, as I mentioned before, these are most likely some kind of reflections of the earliest „images of the world“ as „perceived“ and „remembered“ by life, since the earliest living molecule(the first life). It could also be that these images are in some way expressions of certain

properties of all possible life forms including extraterrestrial, that are not based on RNA/DNA, if such entities exist somewhere in the Universe. Whatever their nature might be, those nonorganic living forms in some way have to relate to their environment, at least by being able to distinguish hot from cold, light from dark.

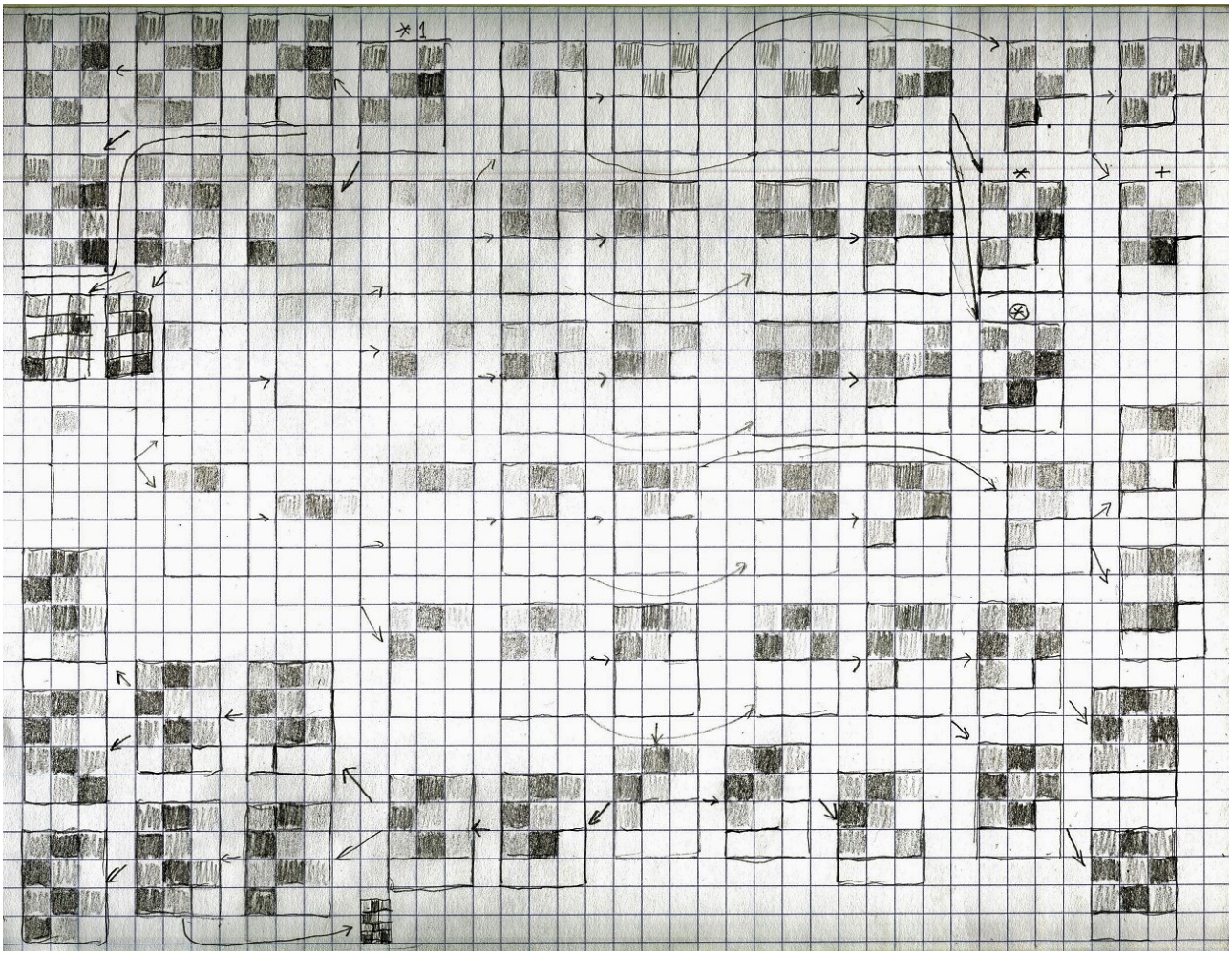
Gregor Mobius 15.12.2021

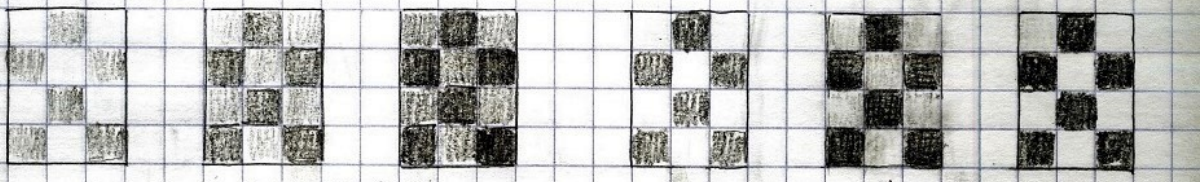
*) *From Chaos to Order and Back,* New York 2017

- - - - -

Apendix 1

Some examples of states of entropy with 2, 3 and 4 bases.





C=0, J=17 D=25%

T	G	T
G	T	G
T	G	T
G	T	G

G	A	G
A	G	A
G	A	G
A	G	A

A	C	A
C	A	C
A	C	A
C	A	C

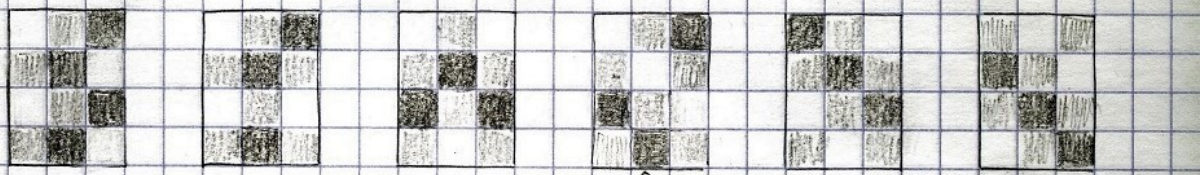
C=0 J=17 D=50%

T	A	T
A	T	A
T	A	T
A	T	A

G	C	G
C	G	C
G	C	G
C	G	C

C=0, J=17 D=75%

T	C	T
C	T	C
T	C	T
C	T	C



T	G	A
G	A	T
T	G	A
G	A	T

T	G	A
G	A	G
T	G	T
G	A	G

T	G	T
G	A	G
A	G	A
G	T	G

o

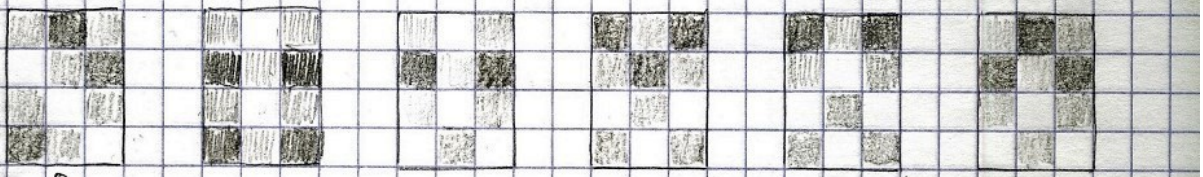
T	G	A
G	T	G
A	G	T
G	A	G

+

A	G	T
G	A	G
T	G	A
G	T	G

+

G	T	G
A	G	T
G	A	G
T	G	A



o

G	A	G
T	G	A
G	T	G
A	G	T

G	T	G
A	G	A
G	T	G
A	G	A

G	T	G
A	G	A
G	T	G
T	G	T

A	G	A
G	A	G
T	G	T
G	T	G

A	G	A
G	T	G
T	G	T
G	A	G

G	A	G
A	G	A
G	T	G
T	G	T

1	2	1
0	1	2
1	0	1
2	1	0

1	0	1
2	1	2
1	0	1
2	1	2

1	0	1
2	1	2
1	0	1
0	1	0

2	1	2
1	2	1
0	1	0
1	0	1

2	1	2
1	0	1
0	1	0
1	2	1

1	2	1
2	1	2
1	0	1
0	1	0

0	1	2
1	2	0
0	1	2
1	2	0

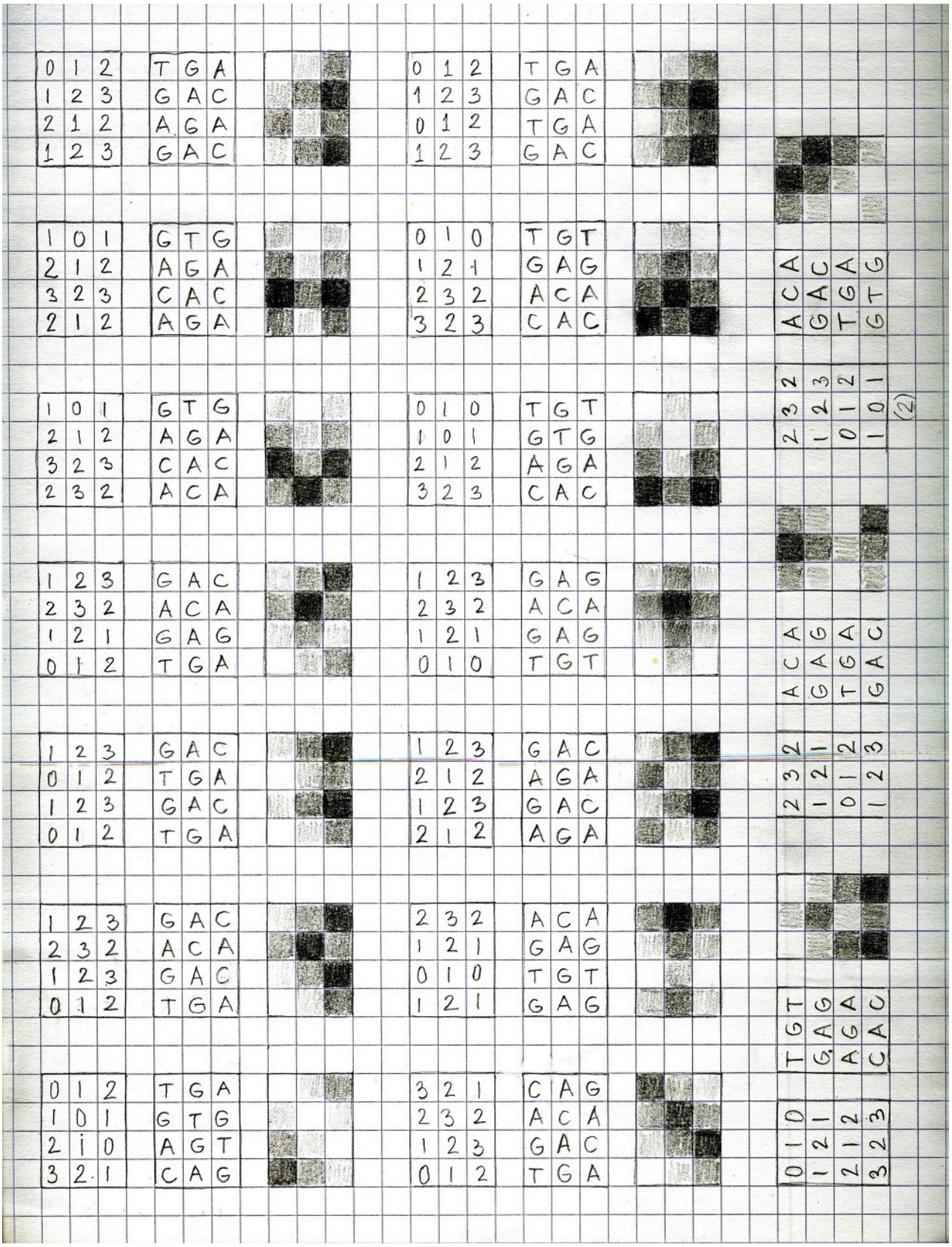
0	1	2
1	2	1
0	1	0
1	2	1

0	1	0
1	2	1
2	1	2
1	0	1

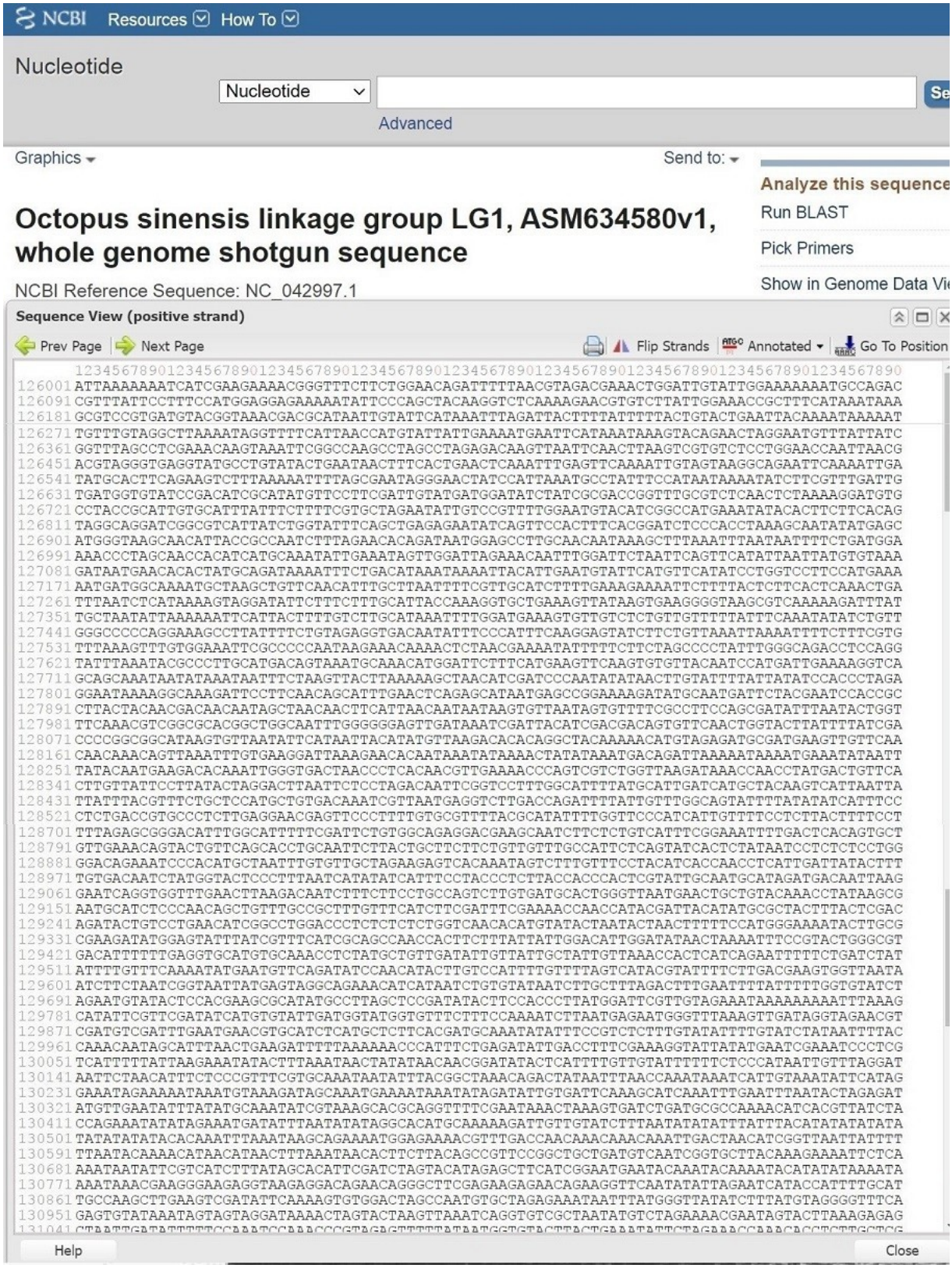
0	1	2
1	0	1
2	1	0
1	2	1

2	1	0
1	2	1
0	1	2
1	0	1

1	0	1
2	1	0
1	2	1
0	1	2


0 1 2 | T G A
1 2 3 | G A C
2 1 2 | A G A
1 2 3 | G A C
0 1 2 | T G A
1 2 3 | G A C
0 1 2 | T G A
1 2 3 | G A C
1 0 1 | G T G
2 1 2 | A G A
3 2 3 | C A C
2 1 2 | A G A
0 1 0 | T G T
1 2 1 | G A G
2 3 2 | A C A
3 2 3 | C A C
A C A
G A C
T G A
G T G
2 3 2
1 2 3
0 1 2
1 0 1
(2)
1 0 1 | G T G
2 1 2 | A G A
3 2 3 | C A C
2 3 2 | A C A
0 1 0 | T G T
1 0 1 | G T G
2 1 2 | A G A
3 2 3 | C A C
1 2 3 | G A C
2 3 2 | A C A
1 2 1 | G A G
0 1 2 | T G A
1 2 3 | G A G
2 3 2 | A C A
1 2 1 | G A G
0 1 0 | T G T
A C A
G A G
T G A
G A C
1 2 3 | G A C
0 1 2 | T G A
1 2 3 | G A C
0 1 2 | T G A
1 2 3 | G A C
2 1 2 | A G A
1 2 3 | G A C
2 1 2 | A G A
2 3 2
1 2 1
0 1 2
1 2 3
1 2 3 | G A C
2 3 2 | A C A
1 2 3 | G A C
0 1 2 | T G A
2 3 2 | A C A
1 2 1 | G A G
0 1 0 | T G T
1 2 1 | G A G
T G T
G A G
A G A
C A C
0 1 2 | T G A
1 0 1 | G T G
2 1 0 | A G T
3 2 1 | C A G
3 2 1 | C A G
2 3 2 | A C A
1 2 3 | G A C
0 1 2 | T G A
0 1 0
1 2 1
2 1 2
3 2 3

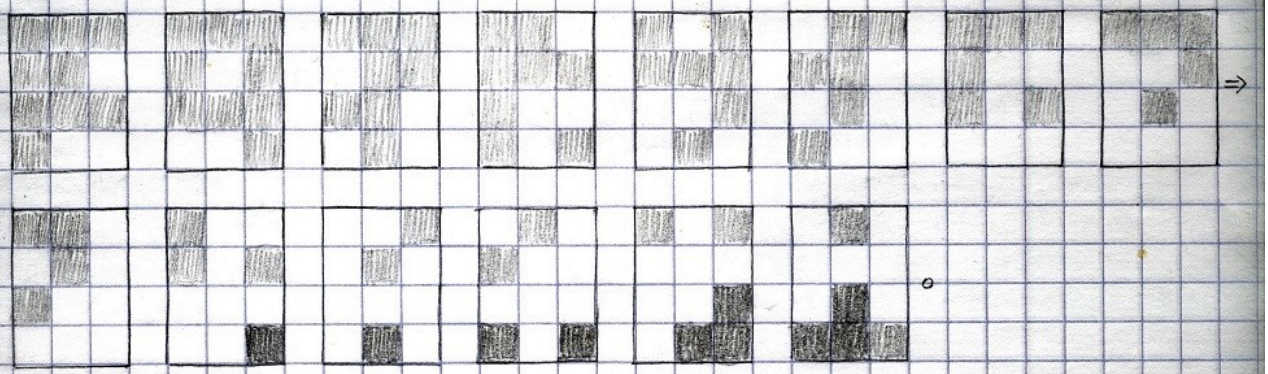
Appendix 2


NCBI Resources How To
Nucleotide
Nucleotide
Advanced
Graphics
Send to:
Analyze this sequence
Run BLAST
Pick Primers
Show in Genome Data Vi
Octopus sinensis linkage group LG1, ASM634580v1, whole genome shotgun sequence
NCBI Reference Sequence: NC_042997.1
Sequence View (positive strand)
Prev Page
Next Page
Flip Strands
Annotated
Go To Position
126001 ATTAAAAAAAATCATCGAAGAAAACGGGTTTCTTCTGGAACAGATTTTTAACGTAGACGAAACTGGATTGTATTGGAAAAAAATGCCAGAC
126091 CGTTTATTCCTTTCCATGGAGGAGAAAAATATTCCCAGCTACAAGGTCTCAAAAGAACGTGTCTTATTGGAAACCGCTTTCATAAATAAA
126181 GCGTCCGTGATGTACGGTAAACGACGCATAATTGTATTCATAAATTTAGATTACTTTTATTTTTACTGTACTGAATTACAAAATAAAAAT
126271 TGTTTGTAGGCTTAAAATAGGTTTTCATTAACCATGTATTATTGAAAATGAATTCATAAATAAAGTACAGAACTAGGAATGTTTATTATC
126361 GGTTTAGCCTCGAAACAAGTAAATTCGGCCAAGCCTAGCCTAGAGACAAGTTAATTCAACTTAAGTCGTGTCTCCTGGAACCAATTAACG
126451 ACGTAGGGTGAGGTATGCCTGTATACTGAATAACTTTCACTGAACTCAAATTTGAGTTCAAAATTGTAGTAAGGCAGAATTCAAAATTGA
126541 TATGCACTTCAGAAGTCTTTAAAAATTTTAGCGAATAGGGAACTATCCATTAAATGCCTATTTCCATAATAAAATATCTTCGTTTGATTG
126631 TGATGGTGTATCCGACATCGCATATGTTCCTTCGATTGTATGATGGATATCTATCGCGACCGGTTTGCGTCTCAACTCTAAAAGGATGTG
126721 CCTACCGCATTGTGCATTTATTTCTTTTCGTGCTAGAATATTGTCCGTTTTGGAATGTACATCGGCCATGAAATATACACTTCTTCACAG
126811 TAGGCAGGATCGGCGTCATTATCTGGTATTTCAGCTGAGAGAATATCAGTTCCACTTTCACGGATCTCCCACCTAAAGCAATATATGAGC
126901 ATGGGTAAGCAACATTACCGCCAATCTTTAGAACACAGATAATGGAGCCTTGCAACAATAAAGCTTTAAATTTAATAATTTTCTGATGGA
126991 AAACCCTAGCAACCACATCATGCAAATATTGAAATAGTTGGATTAGAAACAATTTGGATTCTAATTCAGTTCATATTAATTATGTGTAAA
127081 GATAATGAACACACTATGCAGATAAAATTTCTGACATAAATAAAATTACATTGAATGTATTCATGTTCATATCCTGGTCCTTCCATGAAA
127171 AATGATGGCAAAATGCTAAGCTGTTCAACATTTGCTTAATTTTCGTTGCATCTTTTGAAAGAAAATTCTTTTACTCTTCACTCAAACTGA
127261 TTTAATCTCATAAAAGTAGGATATTCTTTCTTTGCATTACCAAAGGTGCTGAAAGTTATAAGTGAAGGGGTAAGCGTCAAAAAGATTTAT
127351 TGCTAATATTAAAAAATTCATTACTTTTGTCTTGCATAAATTTTGGATGAAAGTGTTGTCTCTGTTGTTTTTATTTCAAATATATCTGTT
127441 GGGCCCCCAGGAAAGCCTTATTTTCTGTAGAGGTGACAATATTTCCCATTTCAAGGAGTATCTTCTGTTAAATTAAAATTTTCTTTCGTG
127531 TTTAAAGTTTGTGGAAATTCGCCCCCAATAAGAAACAAAACTCTAACGAAAATATTTTTCTTCTAGCCCCTATTTGGGCAGACCTCCAGG
127621 TATTTAAATACGCCCTTGCATGACAGTAAATGCAAACATGGATTCTTTCATGAAGTTCAAGTGTGTTACAATCCATGATTGAAAAGGTCA
127711 GCAGCAAATAATATAAATAATTTCTAAGTTACTTAAAAAGCTAACATCGATCCCAATATATAACTTGTATTTTATTATATCCACCCTAGA
127801 GGAATAAAAGGCAAAGATTCCTTCAACAGCATTTGAACTCAGAGCATAATGAGCCGGAAAAGATATGCAATGATTCTACGAATCCACCGC
127891 CTTACTACAACGACAACAATAGCTAACAACTTCATTAACAATAATAAGTGTTAATAGTGTTTTCGCCTTCCAGCGATATTTAATACTGGT
127981 TTCAAACGTCGGCGCACGGCTGGCAATTTGGGGGGAGTTGATAAATCGATTACATCGACGACAGTGTTCAACTGGTACTTATTTTATCGA
128071 CCCCGGCGGCATAAGTGTTAATATTCATAATTACATATGTTAAGACACACAGGCTACAAAAACATGTAGAGATGCGATGAAGTTGTTCAA
128161 CAACAAACAGTTAAATTTGTGAAGGATTAAAGAACACAATAAATATAAAACTATATAAATGACAGATTAAAAATAAAATGAAATATAATT
128251 TATACAATGAAGACACAAATTGGGTGACTAACCCTCACAACGTTGAAAACCCAGTCGTCTGGTTAAGATAAACCAACCTATGACTGTTCA
128341 CTTGTTATTCCTTATACTAGGACTTAATTCTCCTAGACAATTCGGTCCTTTGGCATTTTATGCATTGATCATGCTACAAGTCATTAATTA
128431 TTATTTACGTTTCTGCTCCATGCTGTGACAAATCGTTAATGAGGTCTTGACCAGATTTTATTGTTTGGCAGTATTTTATATATCATTTCC
128521 CTCTGACCGTGCCCTCTTGAGGAACGAGTTCCCTTTTGTGCGTTTTACGCATATTTTGGTTCCCATCATTGTTTTCCTCTTACTTTTCCT
128701 TTTAGAGCGGGACATTTGGCATTTTTCGATTCTGTGGCAGAGGACGAAGCAATCTTCTCTGTCATTTCGGAAATTTTGACTCACAGTGCT
128791 GTTGAAACAGTACTGTTCAGCACCTGCAATTCTTACTGCTTCTTCTGTTGTTTGCCATTCTCAGTATCACTCTATAATCCTCTCTCCTGG
128881 GGACAGAAATCCCACATGCTAATTTGTGTTGCTAGAAGAGTCACAAATAGTCTTTGTTTCCTACATCACCAACCTCATTGATTATACTTT
128971 TGTGACAATCTATGGTACTCCCTTTAATCATATATCATTTCCTACCCTCTTACCACCCACTCGTATTGCAATGCATAGATGACAATTAAG
129061 GAATCAGGTGGTTTGAACTTAAGACAATCTTTCTTCCTGCCAGTCTTGTGATGCACTGGGTTAATGAACTGCTGTACAAACCTATAAGCG
129151 AATGCATCTCCCAACAGCTGTTTGCCGCTTTGTTTCATCTTCGATTTCGAAAACCAACCATACGATTACATATGCGCTACTTTACTCGAC
129241 AGATACTGTCCTGAACATCGGCCTGGACCCTCTCTCTCTGGTCAACACATGTATACTAATACTAACTTTTTCCATGGGAAAATACTTGCG
129331 CGAAGATATGGAGTATTTATCGTTTCATCGCAGCCAACCACTTCTTTATTATTGGACATTGGATATAACTAAAATTTCCGTACTGGGCGT
129421 GACATTTTTTGAGGTGCATGTGCAAACCTCTATGCTGTTGATATTGTTATTGCTATTGTTAAACCACTCATCAGAATTTTTCTGATCTAT
129511 ATTTTGTTTCAAAATATGAATGTTCAGATATCCAACATACTTGTCCATTTTGTTTTAGTCATACGTATTTTCTTGACGAAGTGGTTAATA
129601 ATCTTCTAATCGGTAATTATGAGTAGGCAGAAACATCATAATCTGTGTATAATCTTGCTTTAGACTTTGAATTTTATTTTTGGTGTATCT
129691 AGAATGTATACTCCACGAAGCGCATATGCCTTAGCTCCGATATACTTCCACCCTTATGGATTCGTTGTAGAAATAAAAAAAAATTTAAAG
129781 CATATTCGTTCGATATCATGTGTATTGATGGTATGGTGTTTCTTTCCAAAATCTTAATGAGAATGGGTTTAAAGTTGATAGGTAGAACGT
129871 CGATGTCGATTTGAATGAACGTGCATCTCATGCTCTTCACGATGCAAATATATTTCCGTCTCTTTGTATATTTTGTATCTATAATTTTAC
129961 CAAACAATAGCATTTAACTGAAGATTTTTAAAAAACCCATTTCTGAGATATTGACCTTTCGAAAGGTATTATATGAATCGAAATCCCTCG
130051 TCATTTTTATTAAGAAATATACTTTAAATAACTATATAACAACGGATATACTCATTTTGTTGTATTTTTTCTCCCATAATTGTTTAGGAT
130141 AATTCTAACATTTCTCCCGTTTCGTGCAAATAATATTTACGGCTAAACAGACTATAATTTAACCAAATAAATCATTGTAAATATTCATAG
130231 GAAATAGAAAAATAAATGTAAAGATAGCAAATGAAAATAAATATAGATATTGTGATTCAAAGCATCAAATTTGAATTTAATACTAGAGAT
130321 ATGTTGAATATTTATATGCAAATATCGTAAAGCACGCAGGTTTTCGAATAAACTAAAGTGATCTGATGCGCCAAAACATCACGTTATCTA
130411 CCAGAAATATATAGAAATGATATTTAATATATAGGCACATGCAAAAAGATTGTTGTATCTTTAATATATATTTATTTACATATATATATA
130501 TATATATATACACAAATTTAAATAAGCAGAAAATGGAGAAAACGTTTGACCAACAAACAAACAAATTGACTAACATCGGTTAATTATTTT
130591 TTAATACAAAACATAACATAACTTTAAATAACACTTCTTACAGCCGTTCCGGCTGCTGATGTCAATCGGTGCTTACAAAGAAAATTCTCA
130681 AAATAATATTCGTCATCTTTATAGCACATTCGATCTAGTACATAGAGCTTCATCGGAATGAATACAAATACAAAATACATATATAAAATA
130771 AAATAAACGAAGGGAAGAGGTAAGAGGACAGAACAGGGCTTCGAGAAGAGAACAGAAGGTTCAATATATTAGAATCATACCATTTTGCAT
130861 TGCCAAGCTTGAAGTCGATATTCAAAAGTGTGGACTAGCCAATGTGCTAGAGAAATAATTTATGGGTTATATCTTTATGTAGGGGTTTCA
130951 GAGTGTATAAATAGTAGTAGGATAAAACTAGTACTAAGTTAAATCAGGTGTCGCTAATATGTCTAGAAAACGAATAGTACTTAAAGAGAG
Help
Close

OCTOPUS SINESIS LINKAGE GROUP LG1, ASM634580v1

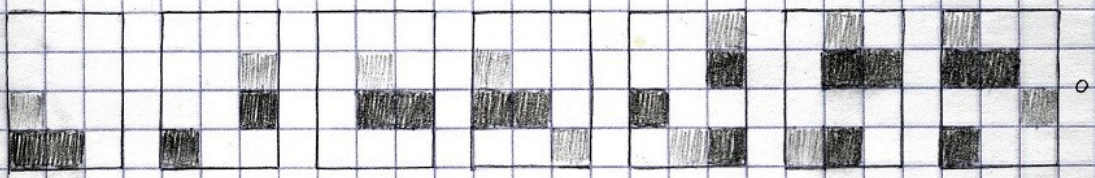
P. 310

A A A A A T A A A A T T A T A T T T T T C T C C A



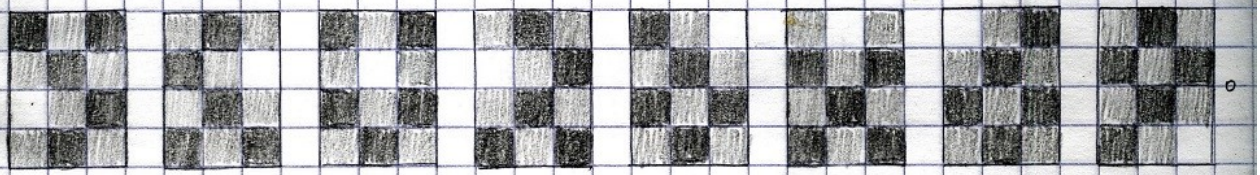
P. 497

T T T T T T A T T C C T T T A C T T

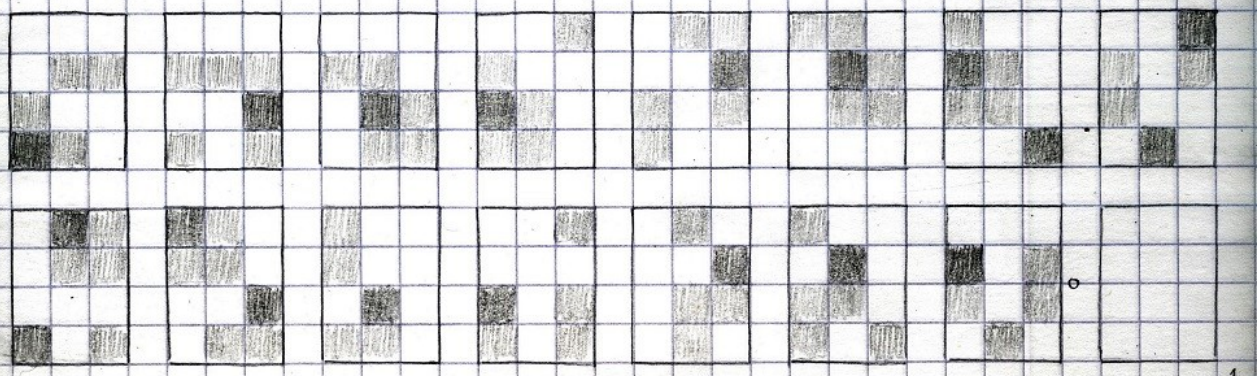


P. 681

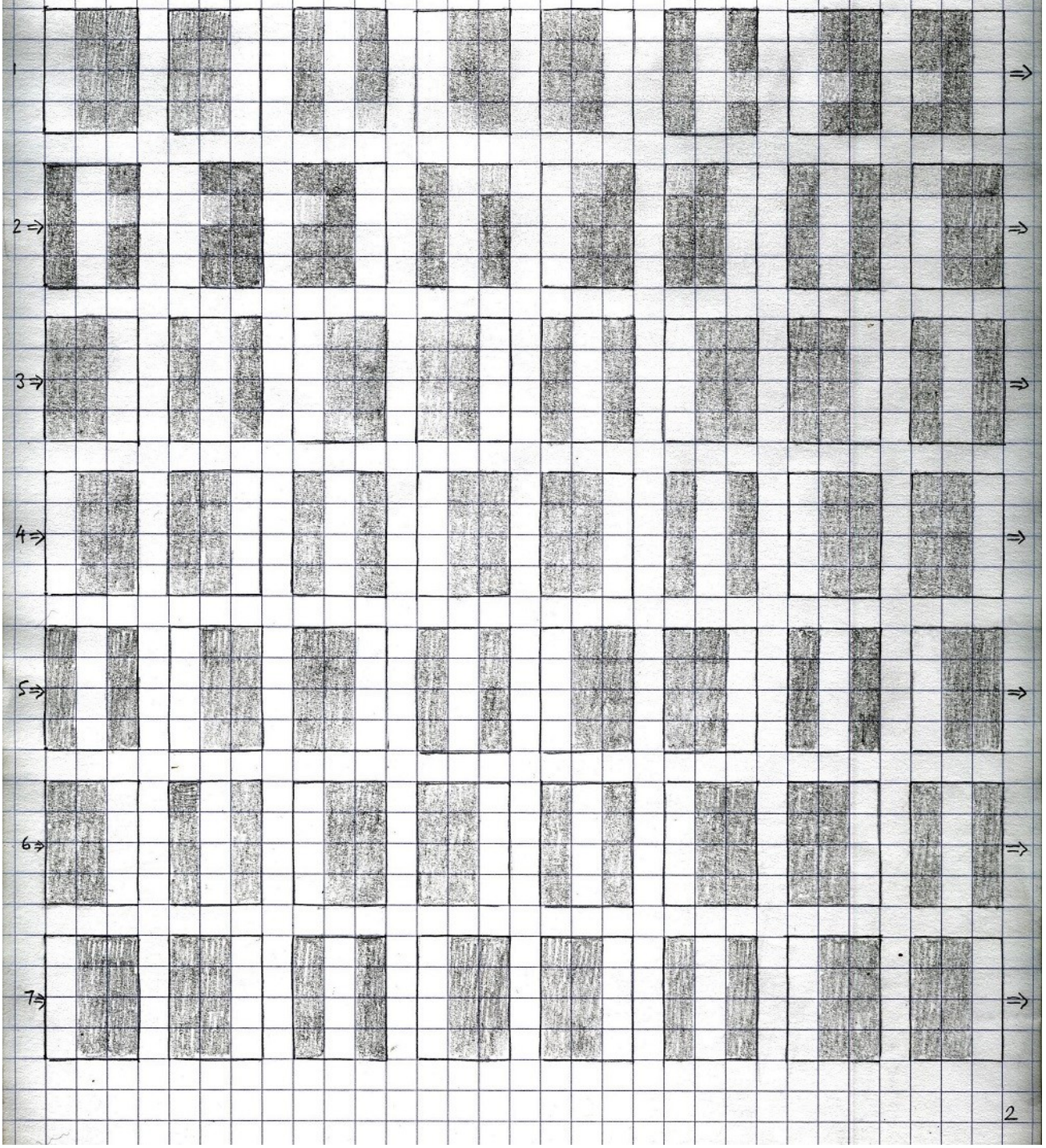
C A C A C A T A C A C A C A C A C A T

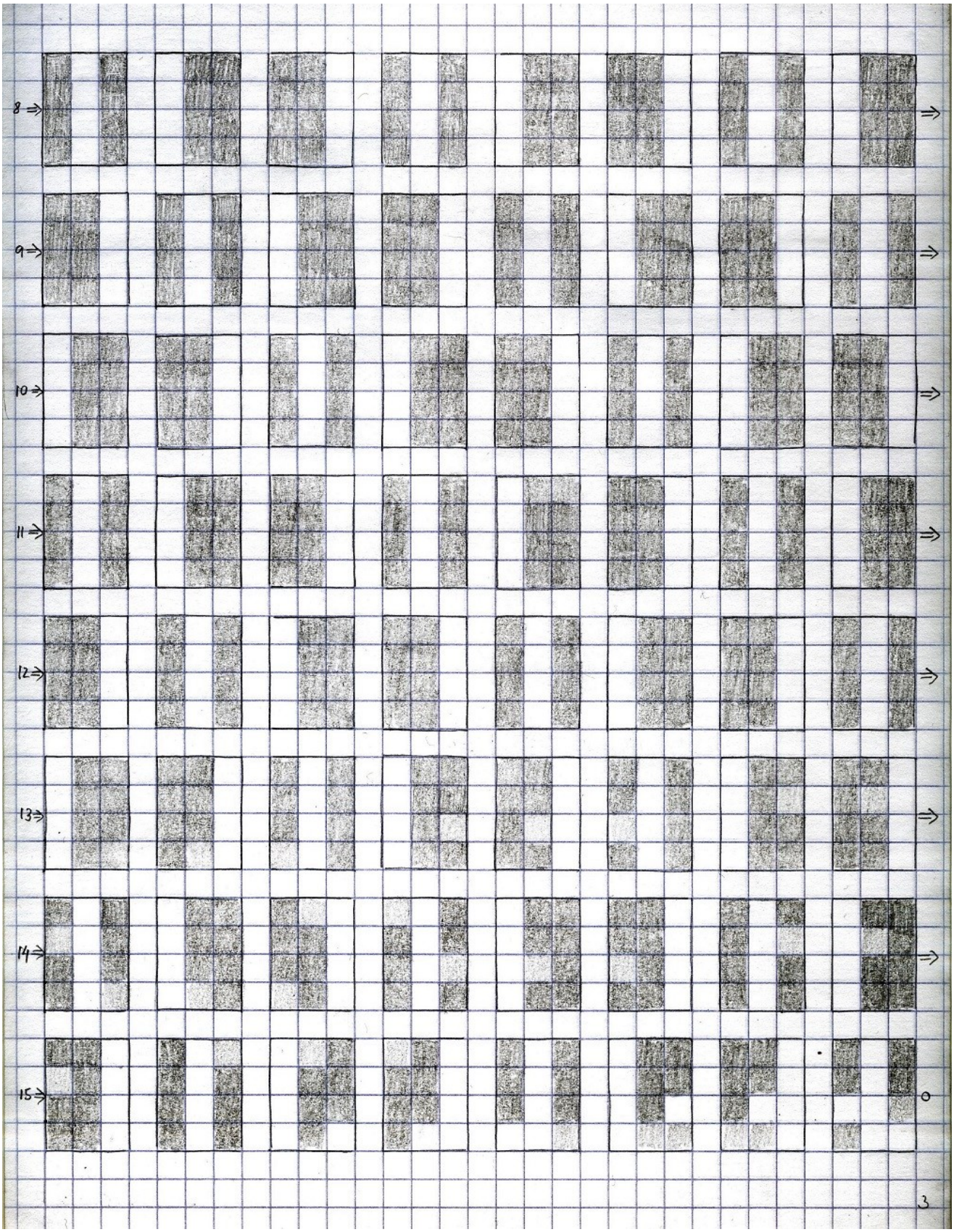


P. 999 T T T T A A A T T C A T A A T T T C T A A T A T A T

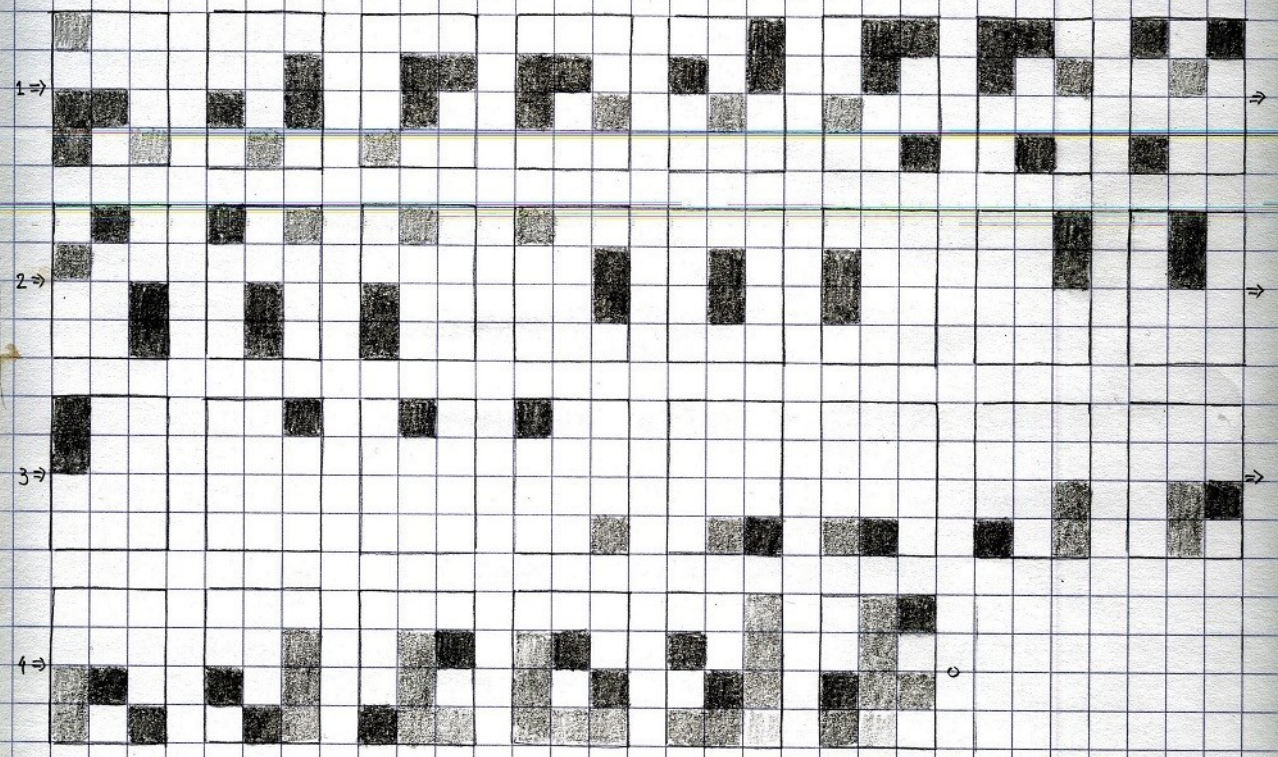


P.1411 TAATAATAATAATGATAATAATAATAATAA
TAATAATA ATAATAAT AATAATAATAATAAT
AATAATAATAATAATAATAATAATAATAATAA
ATAATAATAATAATAATAATGTAATAATGATAA
TAATATTGGTT

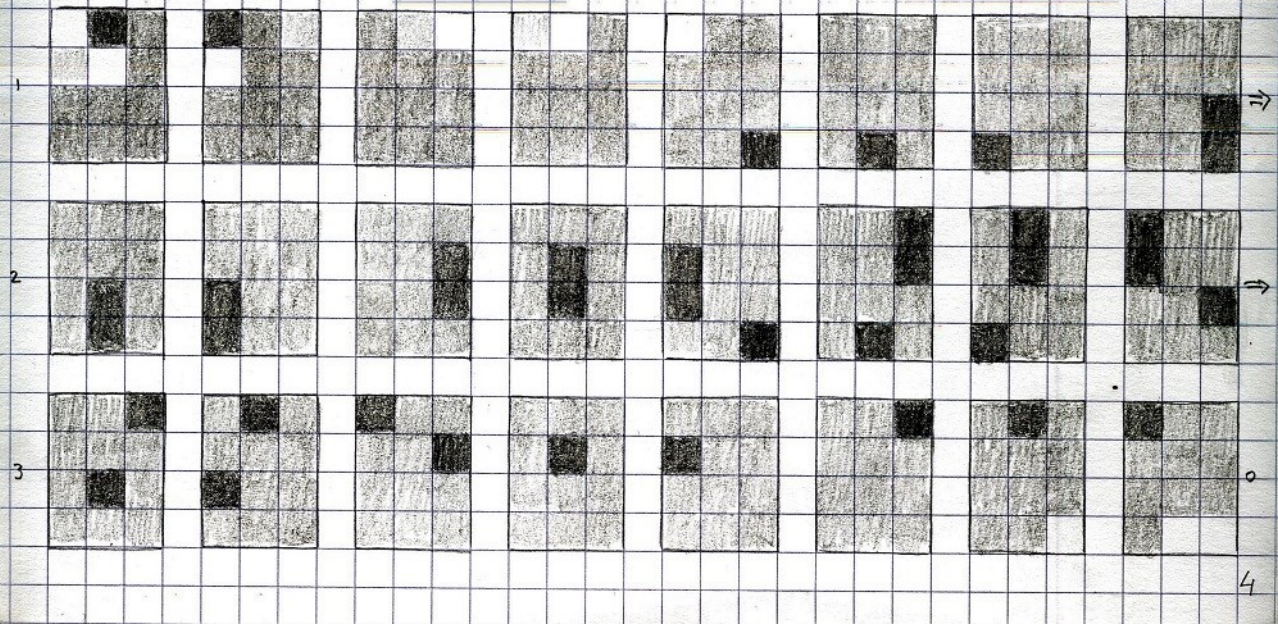



8
9
10
11
12
13
14
15

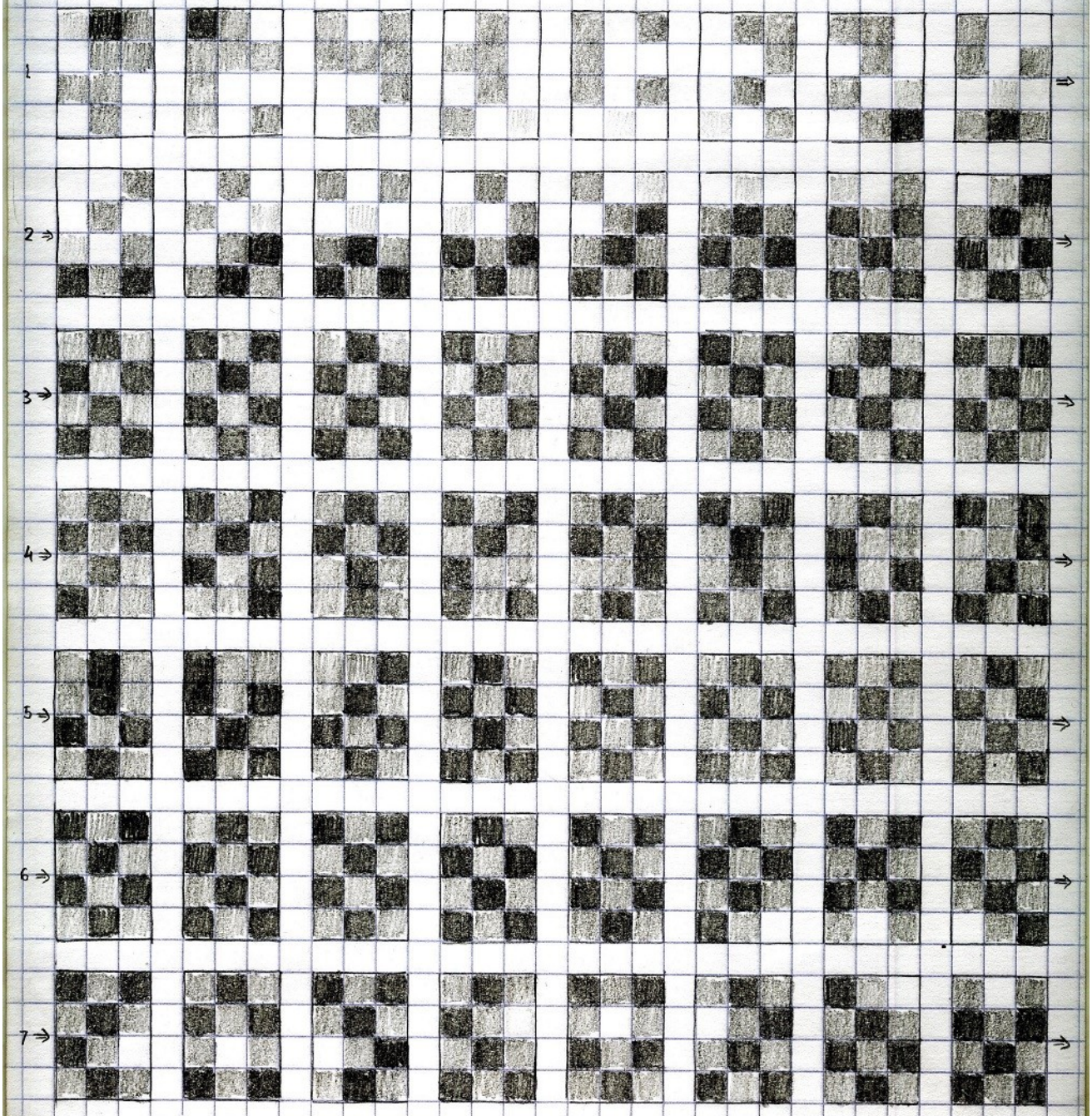
P.4025: A T T T T T C C T C T A T T T T C T T C T T T T T T T T
T T T A C T A T C A A A G T

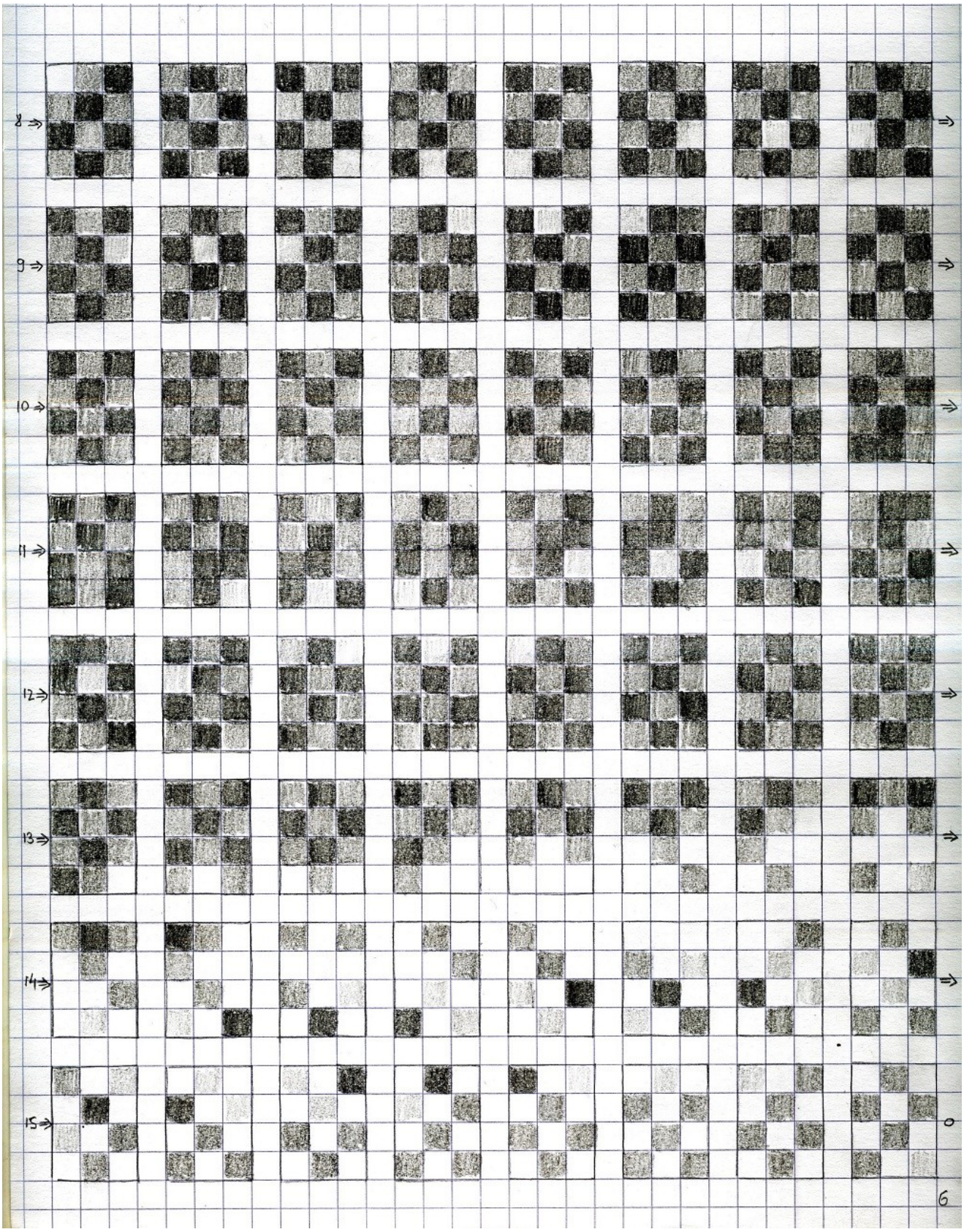


P.4091: T C A G T A A A A A A A A A A A C A A C A A A A C A A A A
A A A A A A T T T

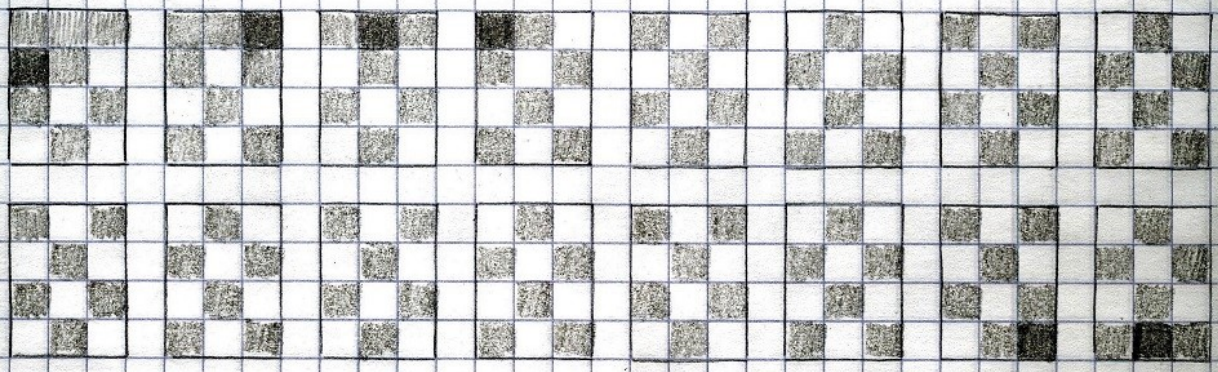


P.4248: GCATAAAATTATATGTACA[1]CACACACA[2]
CACACACA[3]ACACACACAC[4]ACACACACAC[5]ACACAT
AC[6]ACACACACAC[7]ACGCACAC[8]ACACACACAC[9]ACAC
ACCA[10]CGCACACA[11]CACACACA[12]TATTTATG[13]TC
TGTATA[14]TATATATG[15]

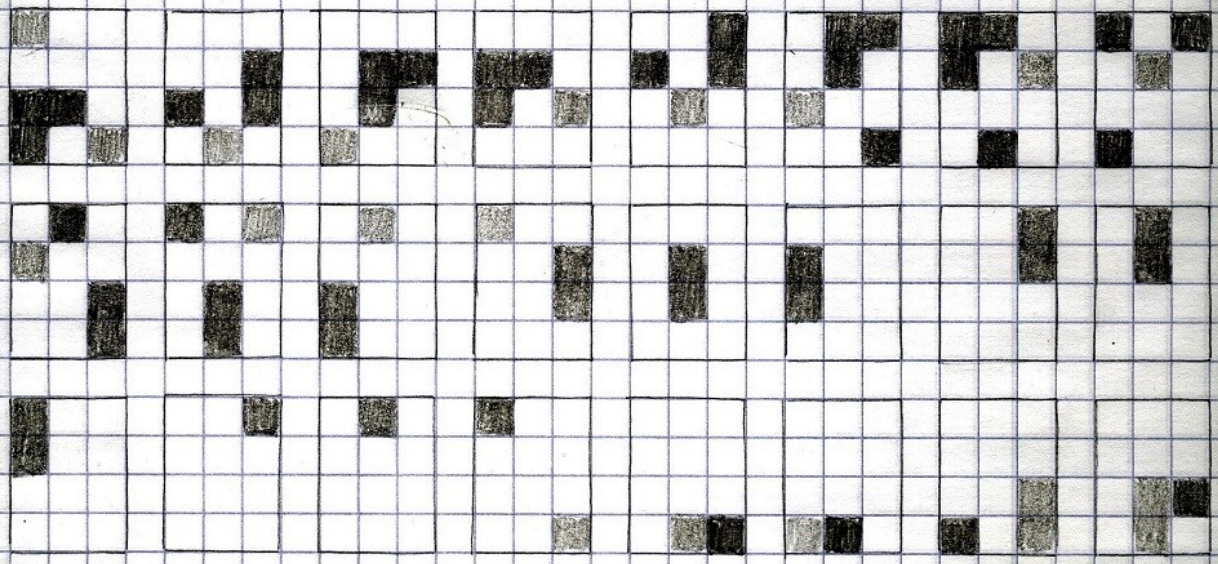



8 ⇒
9 ⇒
10 ⇒
11 ⇒
12 ⇒
13 ⇒
14 ⇒
15 ⇒
⇒
⇒
⇒
⇒
⇒
⇒
⇒
o
6

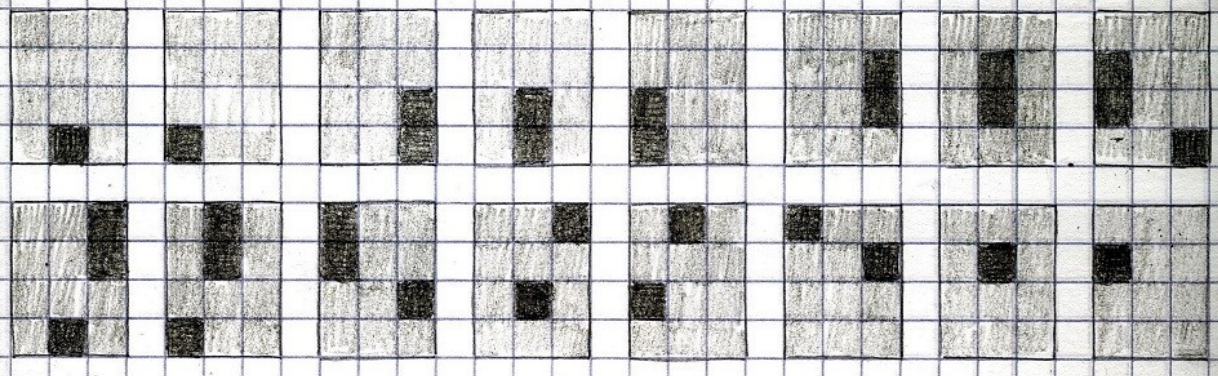
P.4471: A A A C A T A T A T A T A T A T A T A T A T A C A^{2}



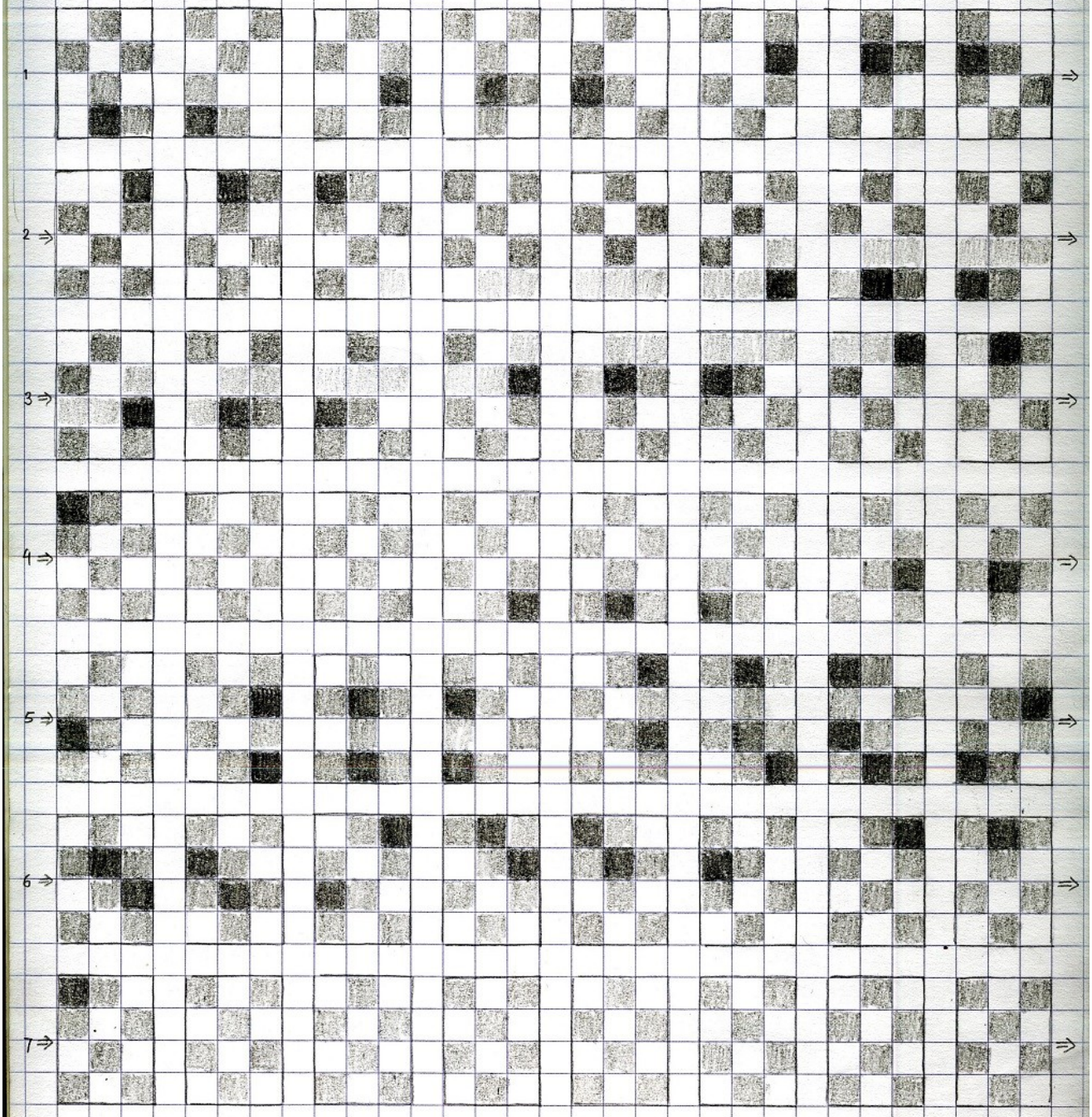
P. 4025: A T T T T T C C T C T A T T T T C T T C T T T T T T T^{2}
T T T A C T A T^{3}

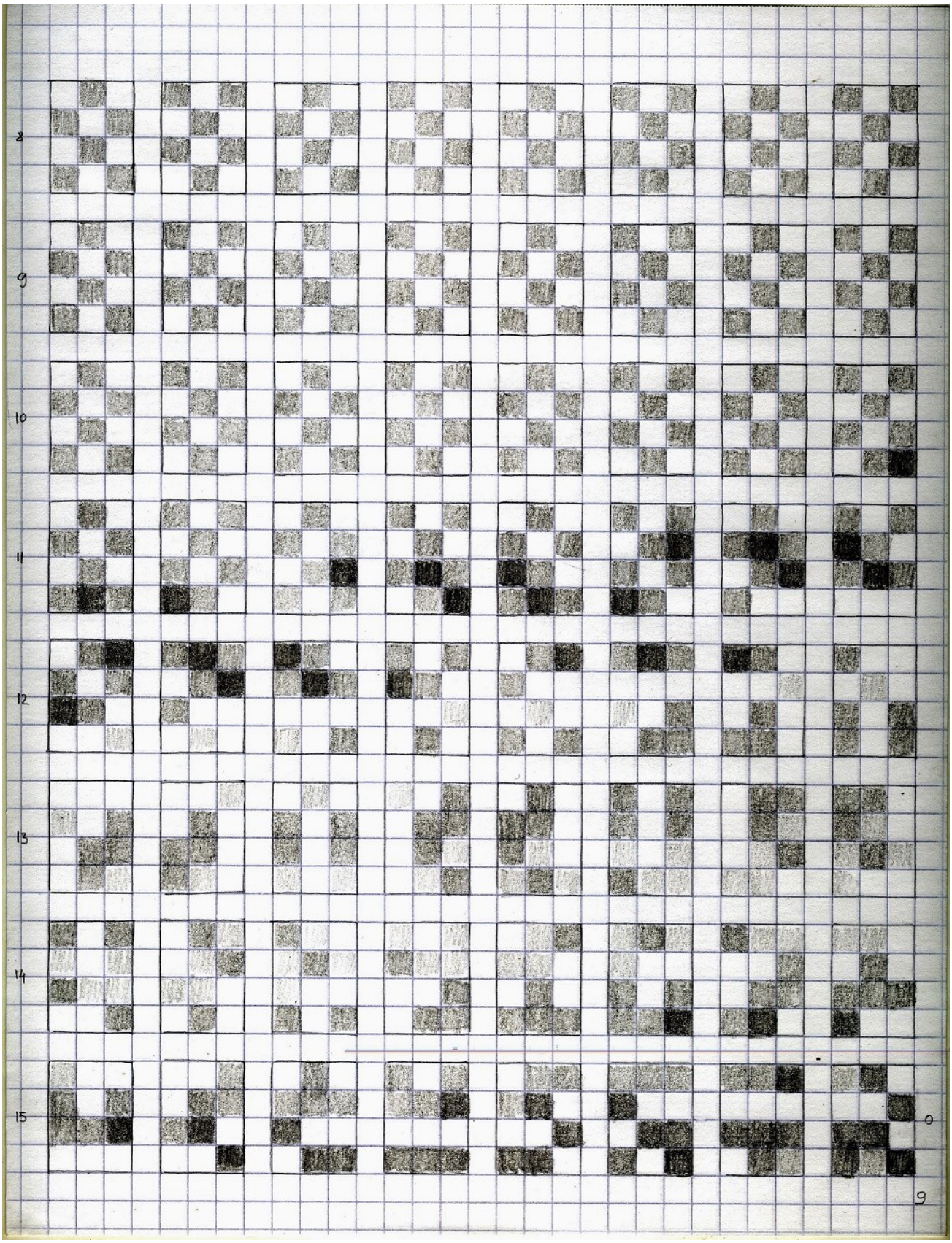


P. 4096: A A A A A A A A A A C A A C A A A A C A A A A A A A A

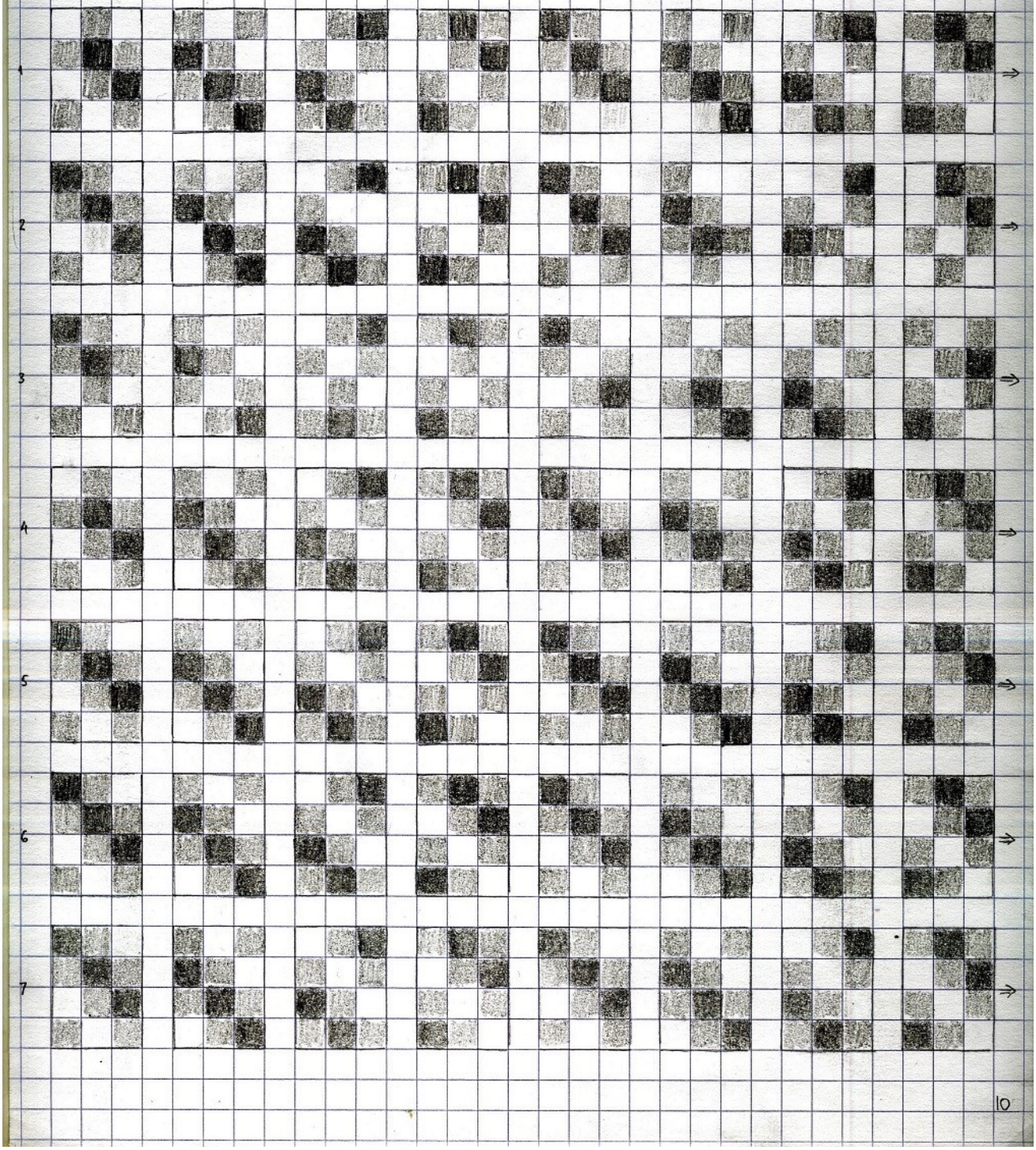


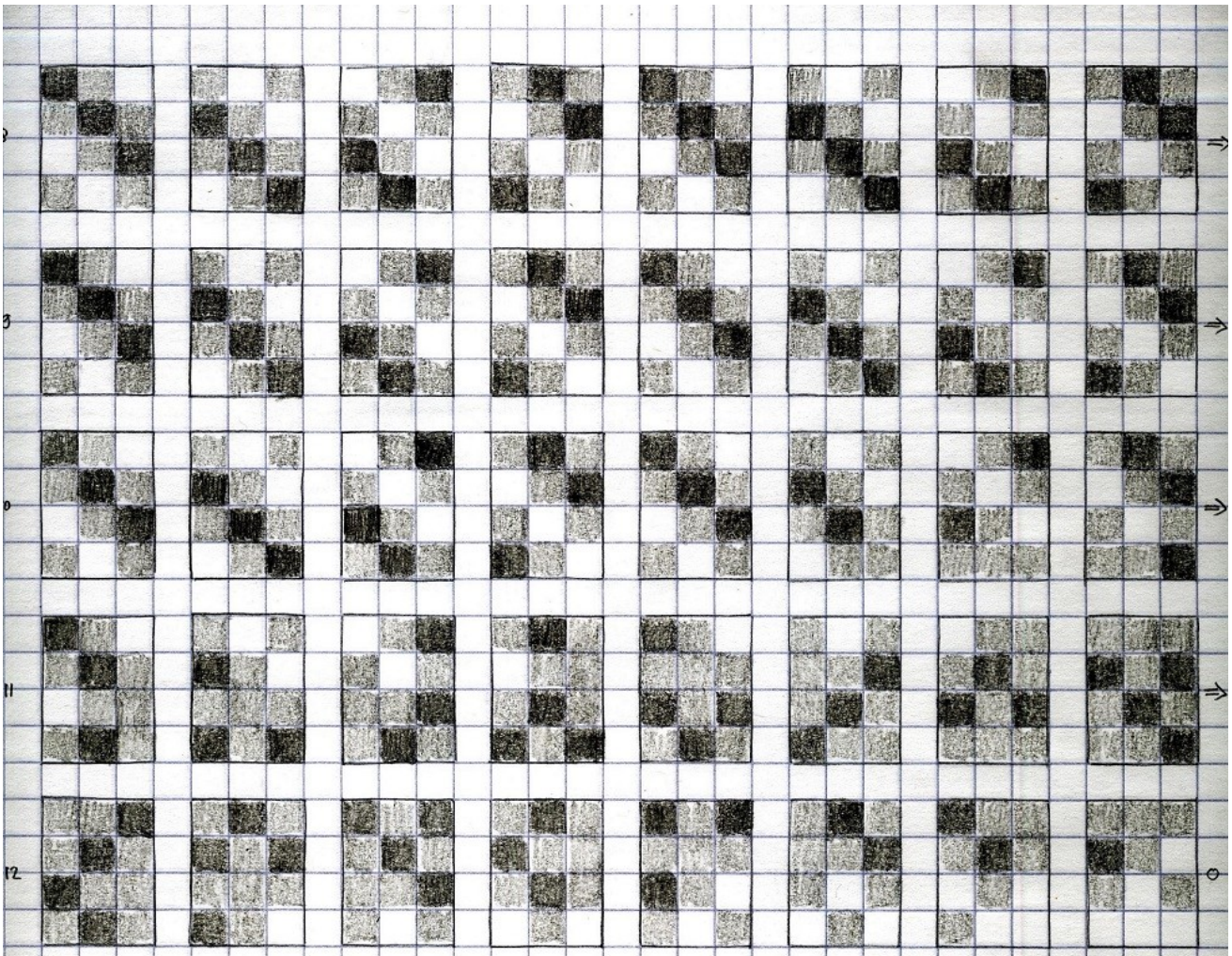
P.4865: TATATATTCATATATATGGGCAT[2]
ATATATAT[3]ATACATAT[4]ACATACAT[5]ATATAT
AT[6]ATATATAT[7]ATATATAT[8]ATATATAT[9]ATAT
ATAC[10]ATACATTT[11]GTAATAATA[12]GTGAGGTT[13]AT
AAACTT[14]TCCCTCAC[15]



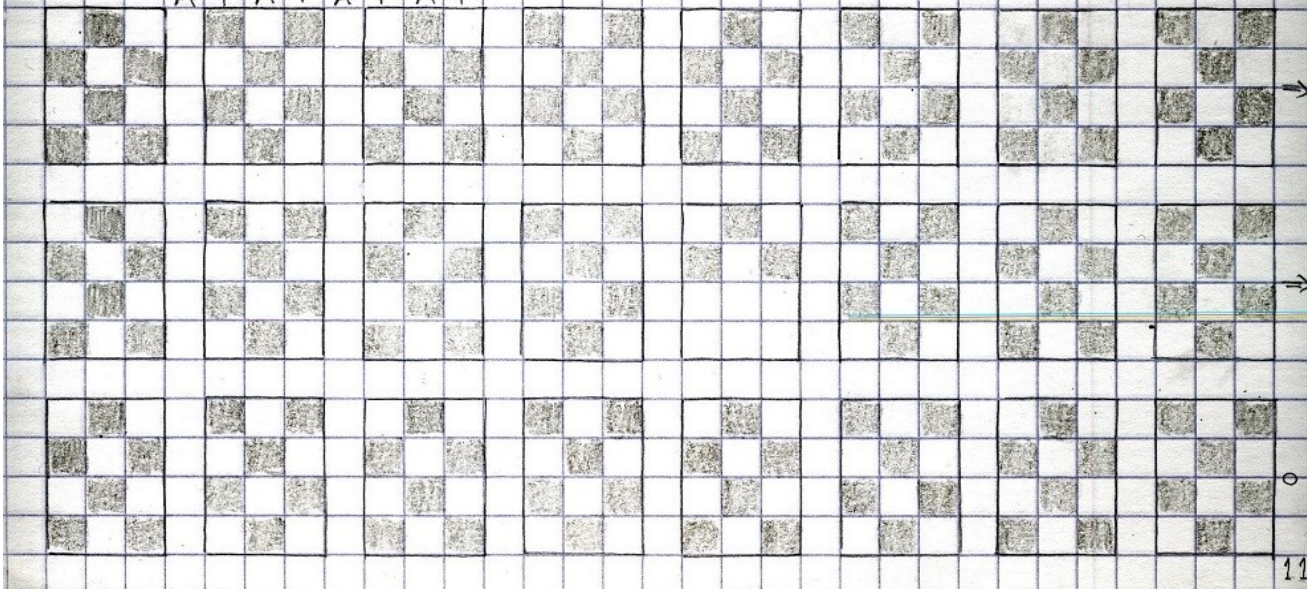

8
9
10
11
12
13
14
15
0
9

P.9043: TATACATACATACATGCAT[1]ACATATAT[2]
ACATACAT[3]ACATACAT[4]ACATACAT[5]ACATAC
AT[6]ACATACAT[7]ACATACAT[8]ACATACAT[9]ACAT
AAAC[10]ACACAAAC[11]ATATATTT[12]



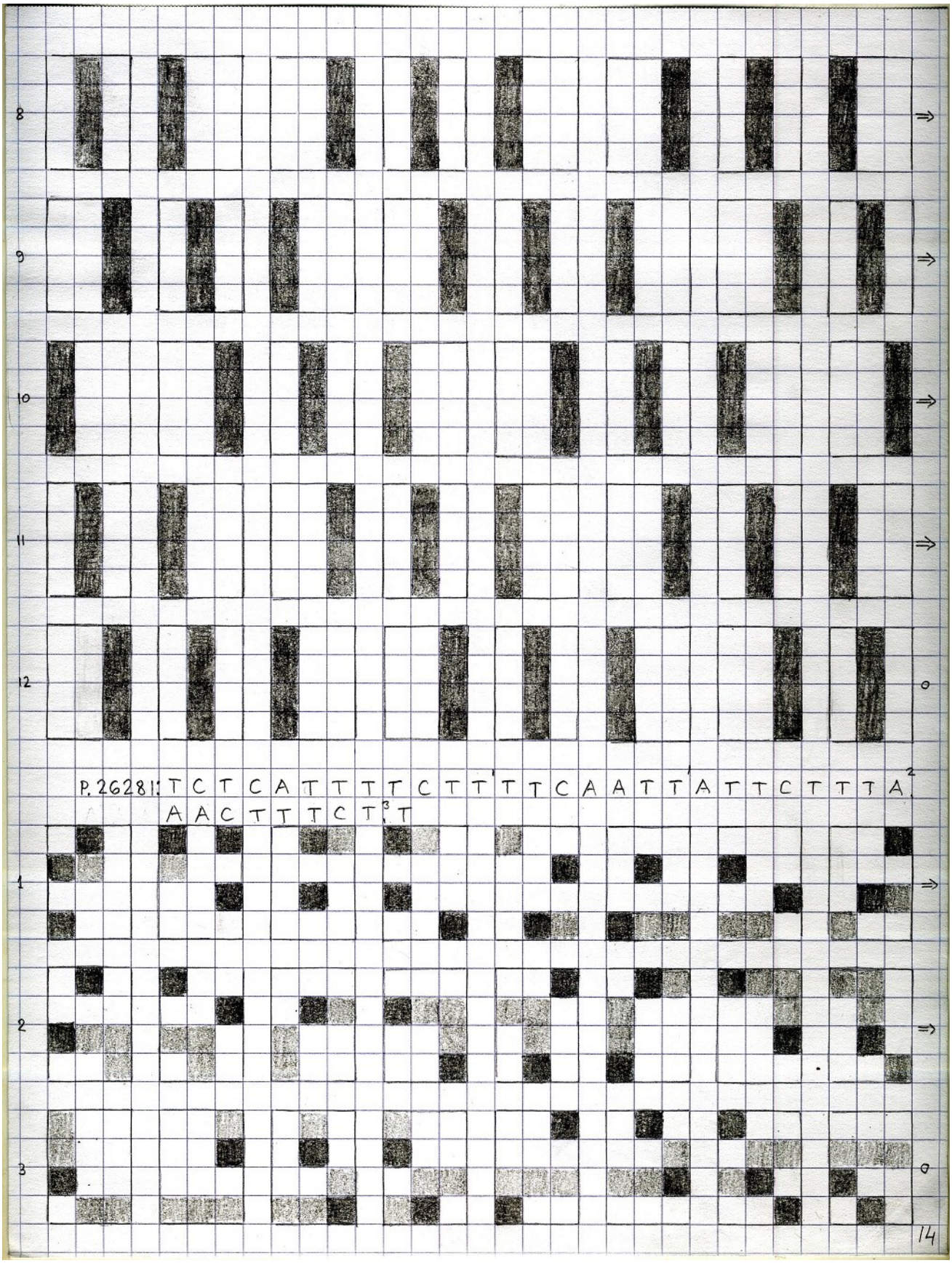


P.14876: TATATATATATATATATATATATATATAT^2
ATATATAT^3

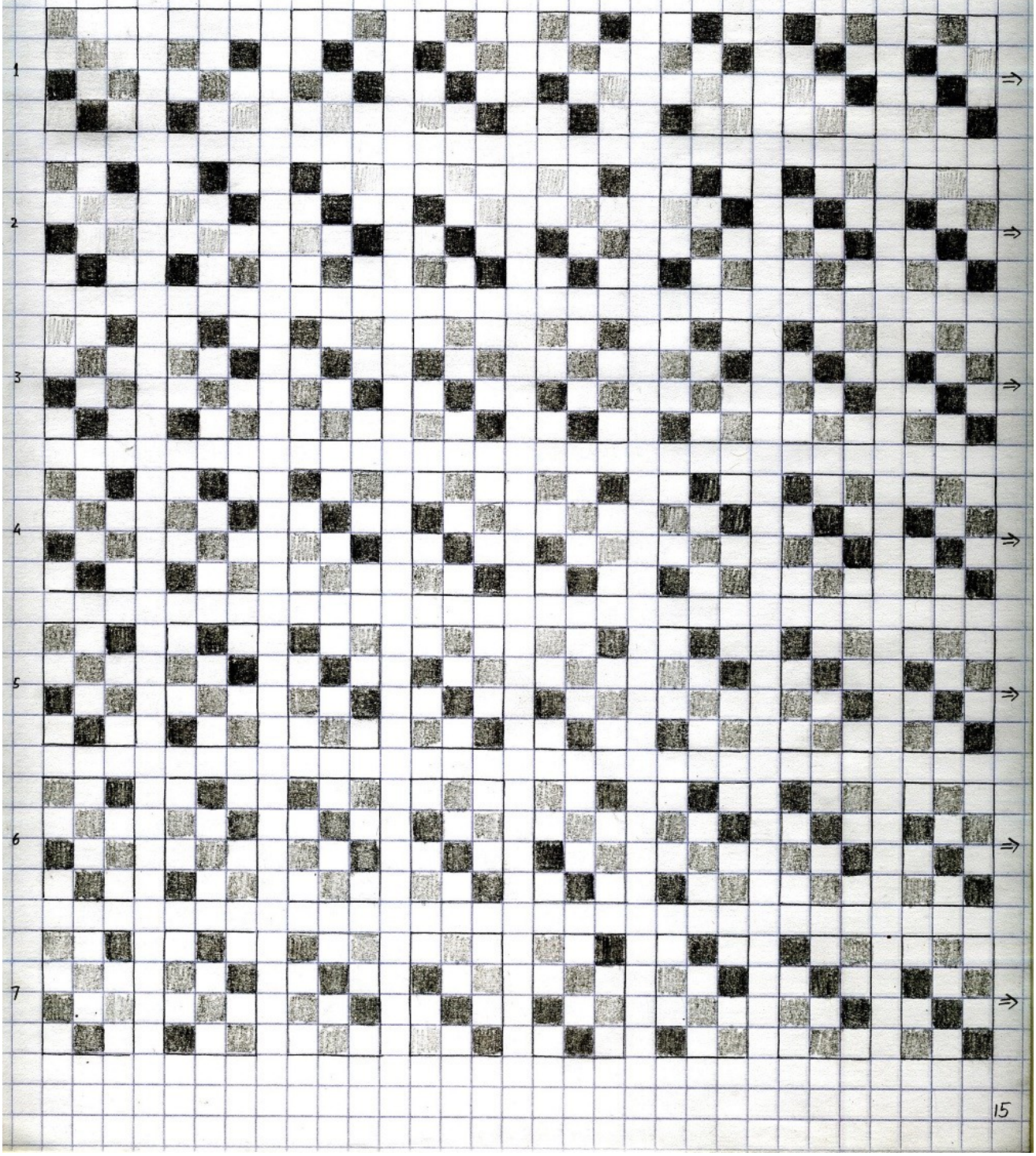


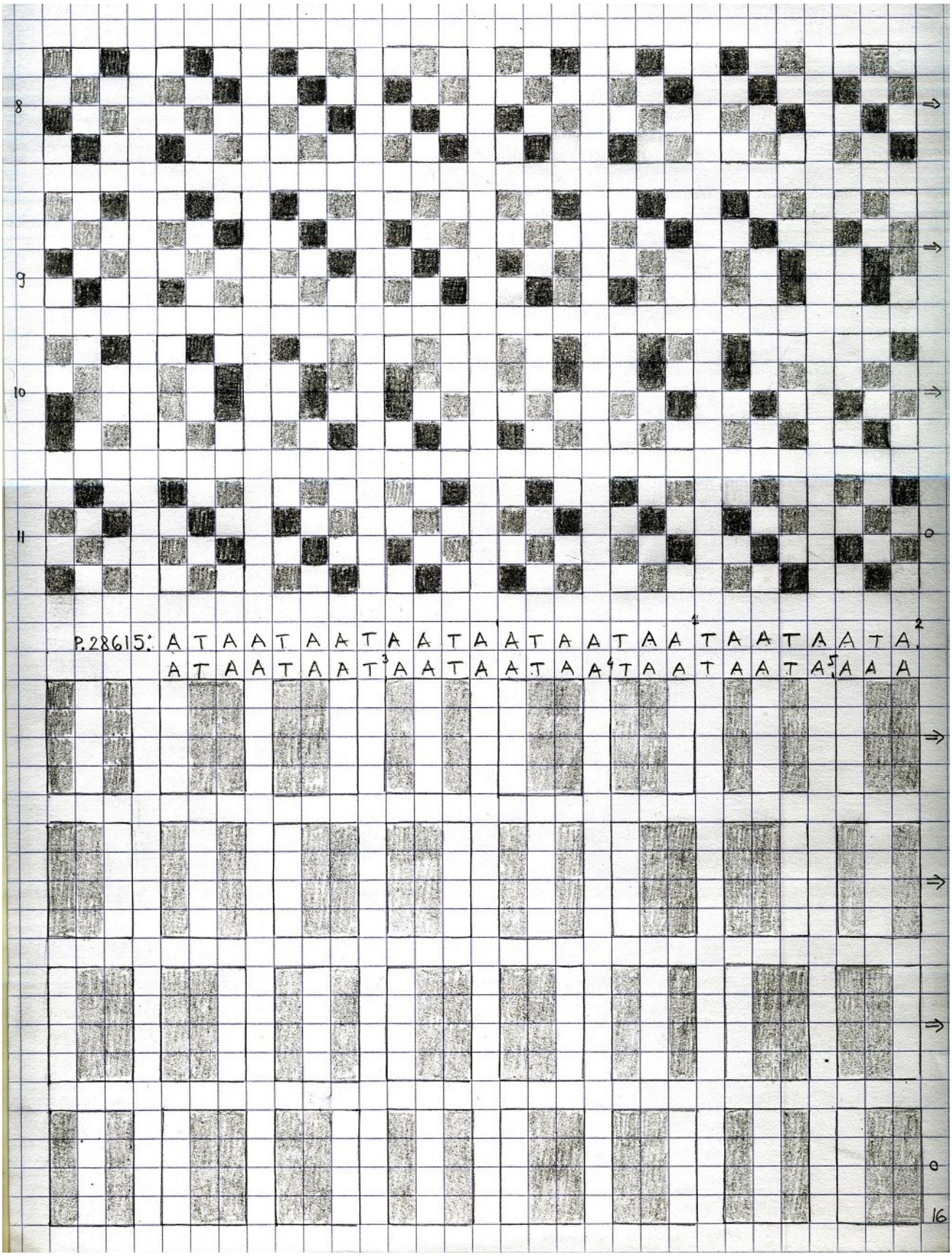
P.1541: A A C A G T G T T T

P. 26493: C T T C A T T T T A C C A T C T T C C T T C T T C T T C C T C T T C T T C T

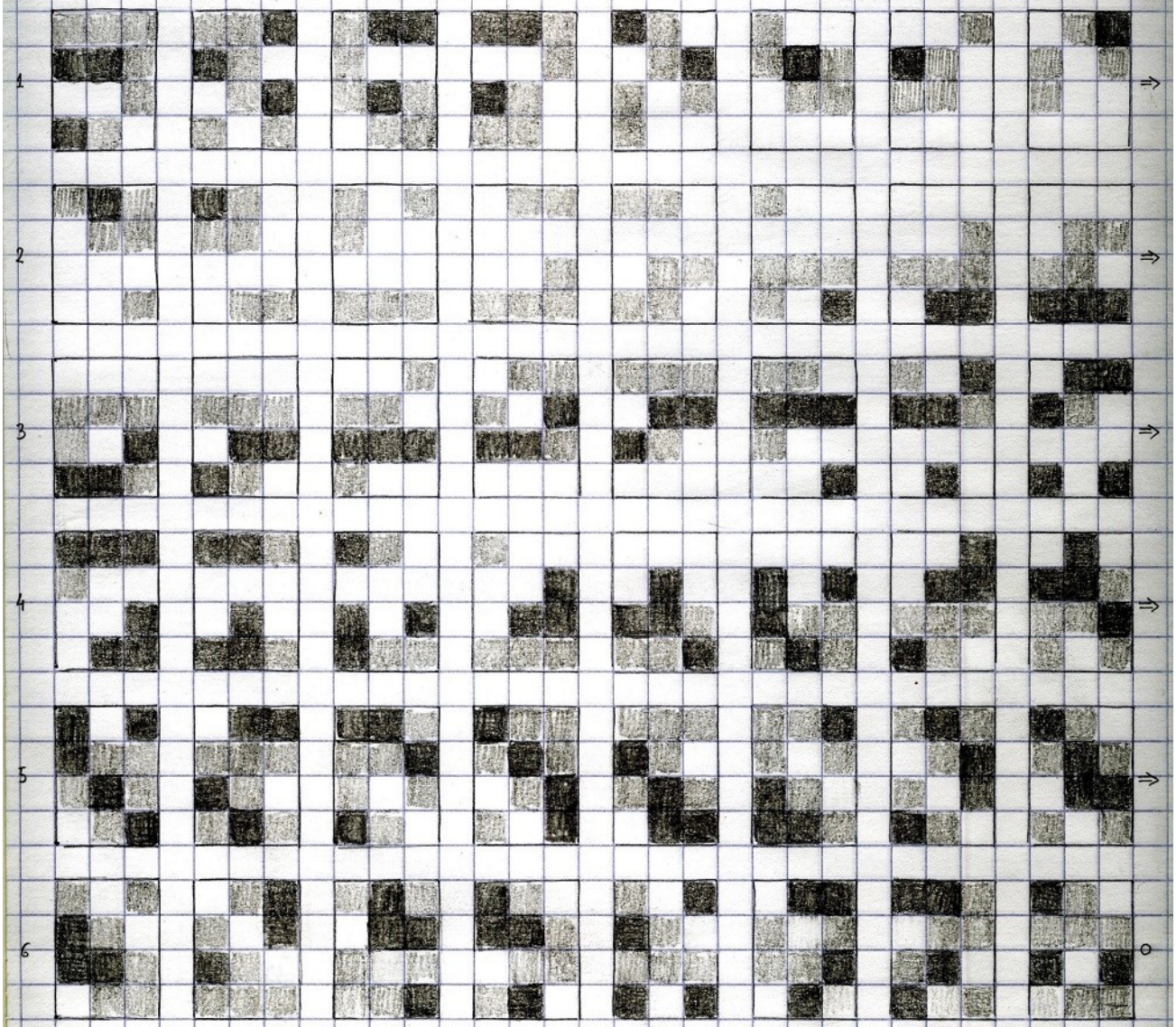

8
9
10
11
12
P. 26281: T C T C A T T T T C T T T T C A A T T A T T C T T T A
A A C T T T C T T
1
2
3

P.26618: ATTTATCTATCTGTCTGTCTATCTATC
TATCTATCTATCTATCTATCTATCTATCTA
TCTATCTATCTATCTATCATCTATCT
ATCTATCTATCT

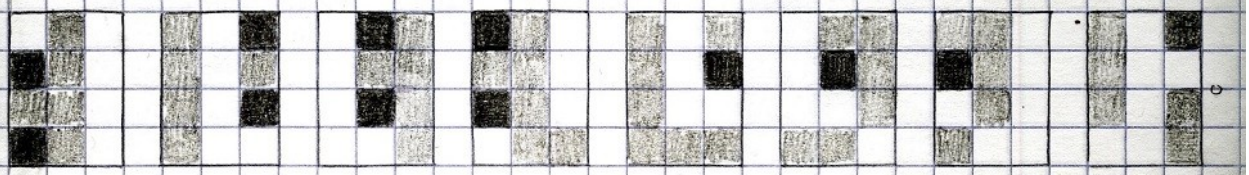



8
9
10
11
P.28615: ATAATAATAATAATAATAATAATAATAATAATA
ATAATAAT AATAATAATA TAATAATAATA AAA
16

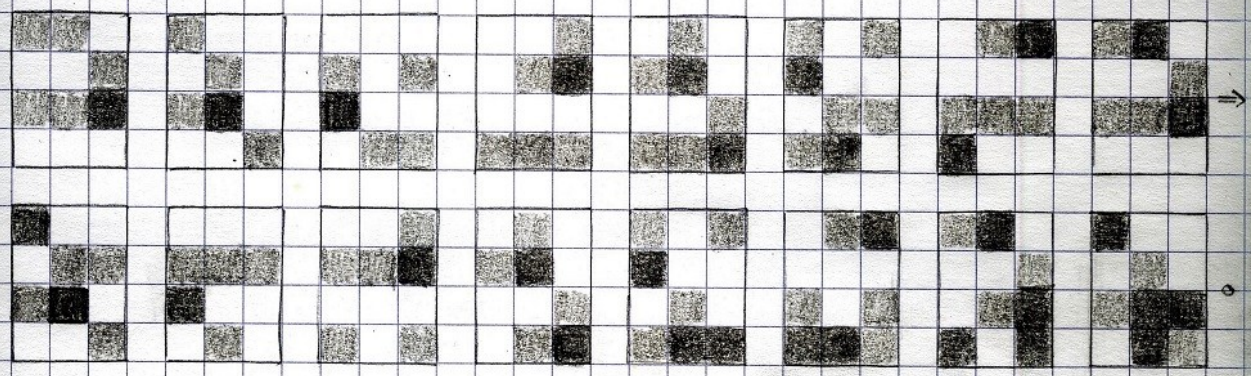
P.32845: A A A C C A T T A C A T A A T T T T T A A A A T C C C
A T T T T C T C C A A A C A T A C A T C C A T A A A C T C G A A



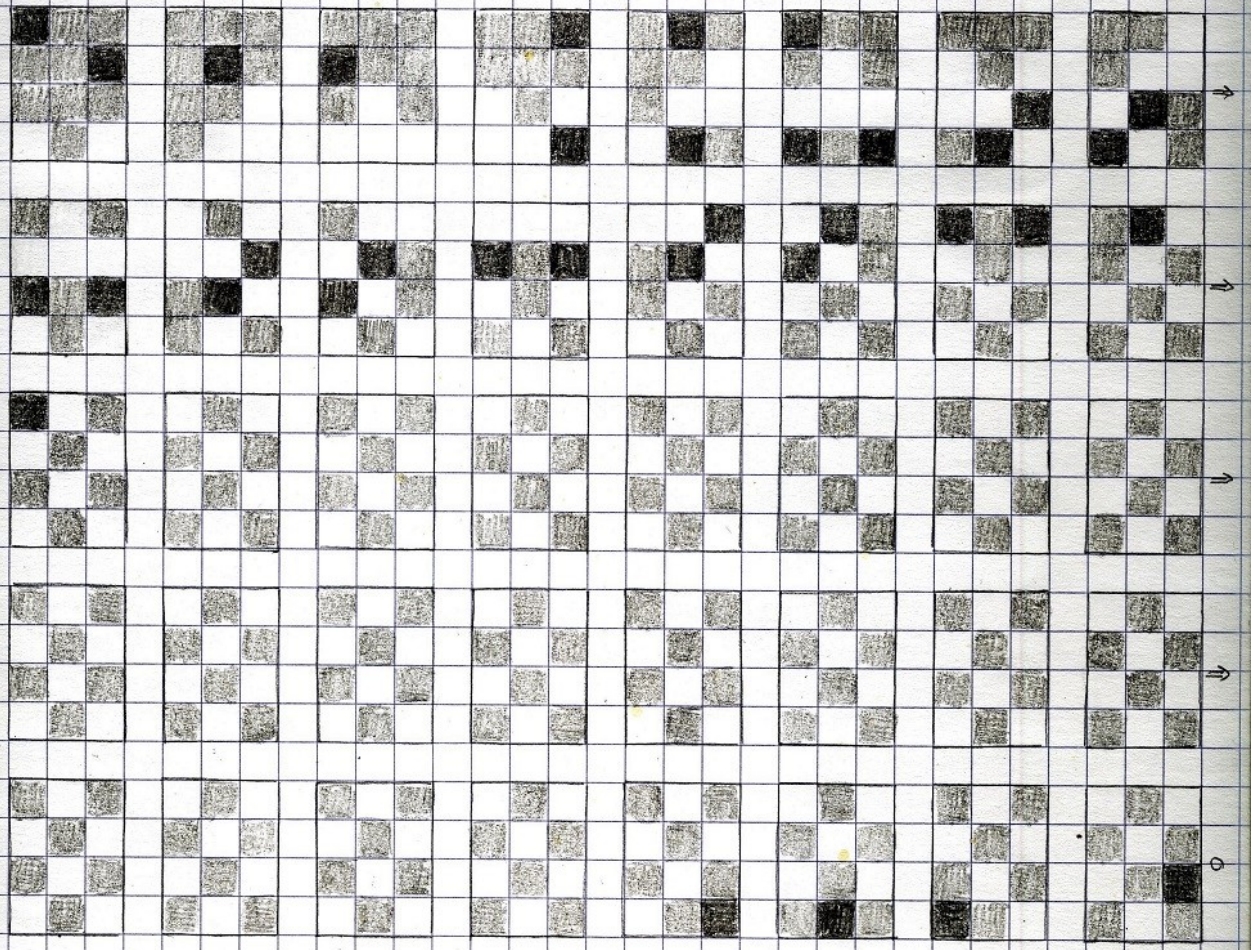
P.33105: T A T C A T A A T C A T T A A A T T A



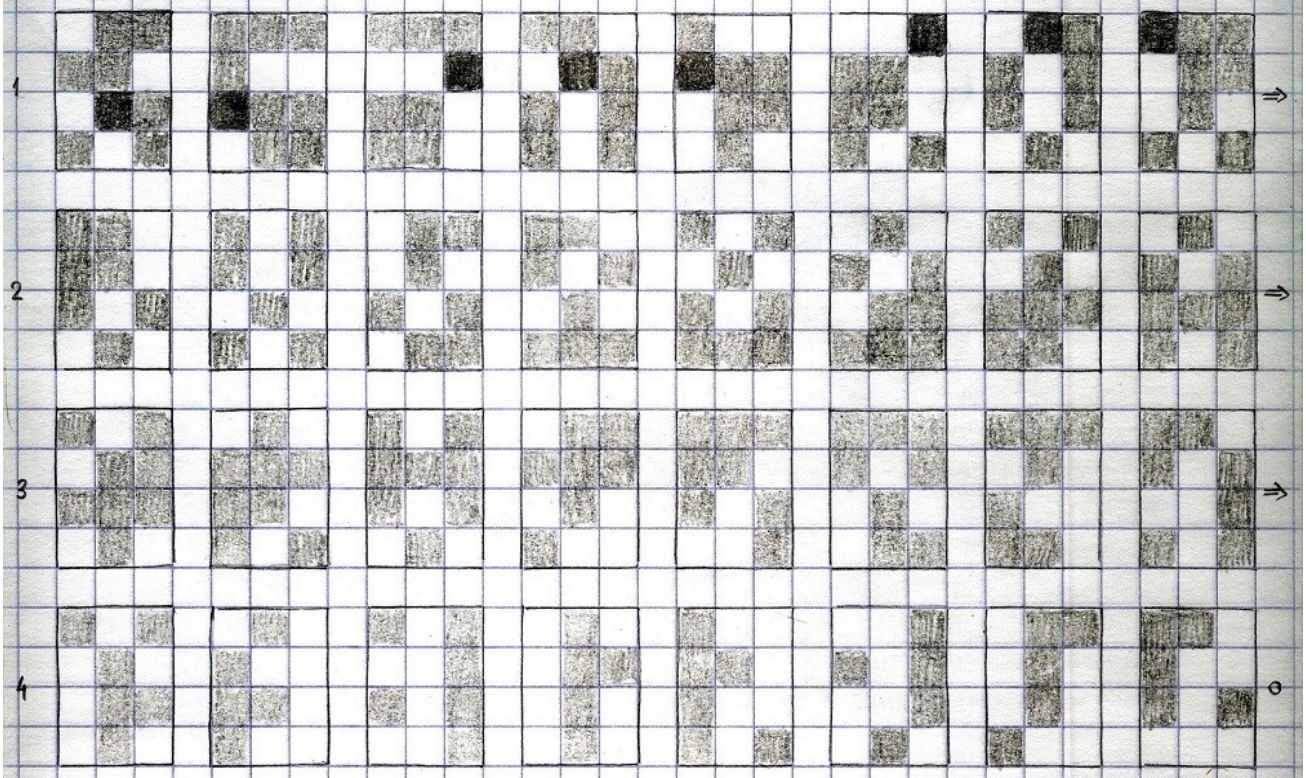
P.33350: AATTTAAACTTTAAACTTTATACCACA



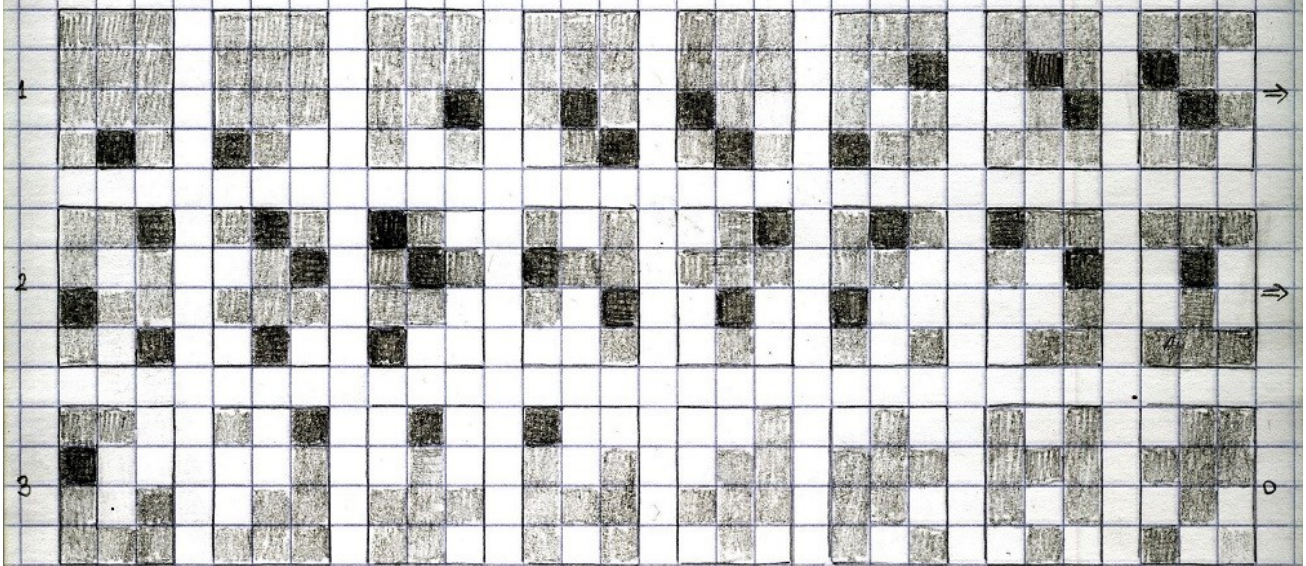
P.33351: CAAAACAAATATTTCACTATATATATA
TATATATA[3]TATATATA[4]TATACATA[5]



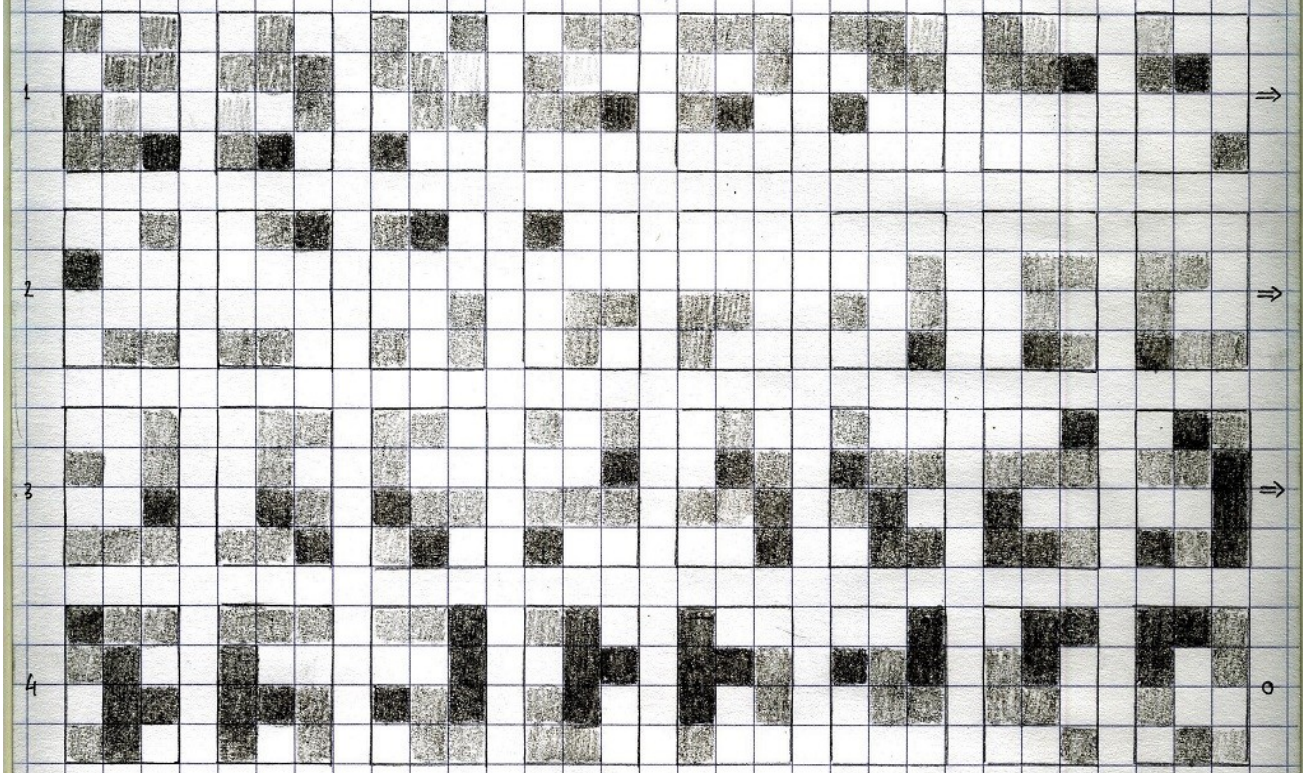
P.34310: TAAAATTCATATATATATATAAAAAATA
TATTAATATTATATTT



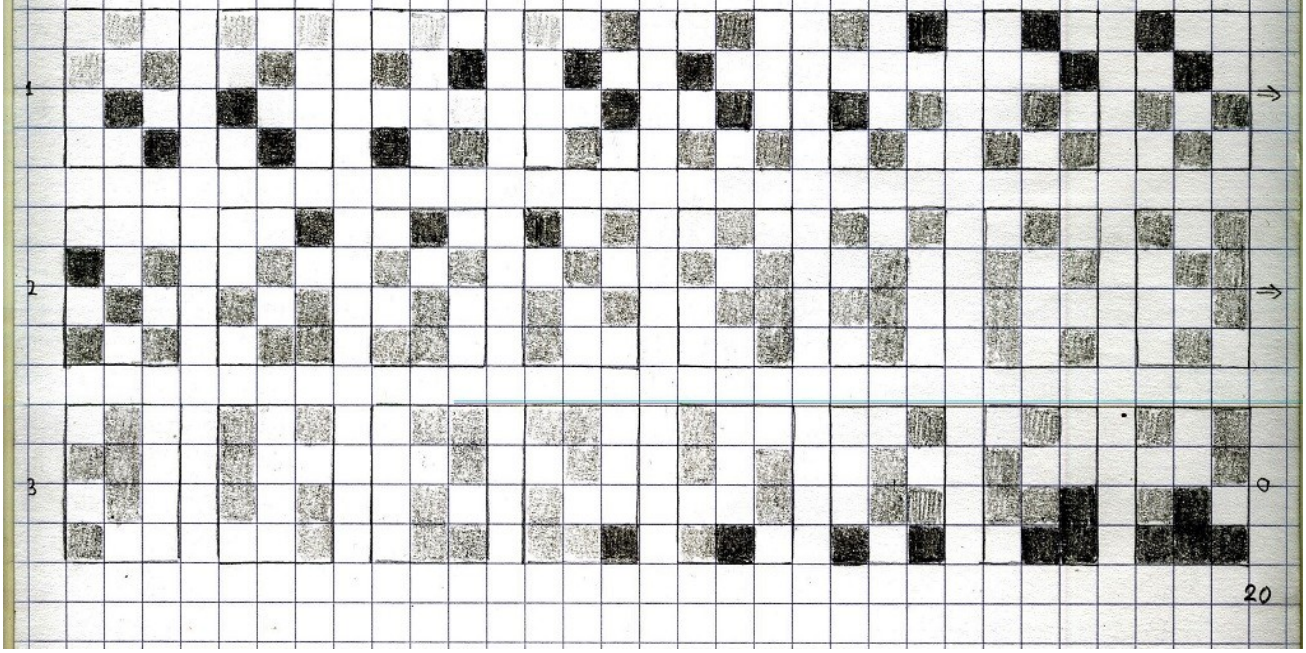
P.34354: AAAAAAAAAAACATACAAATCTTATAAA
AATATATG



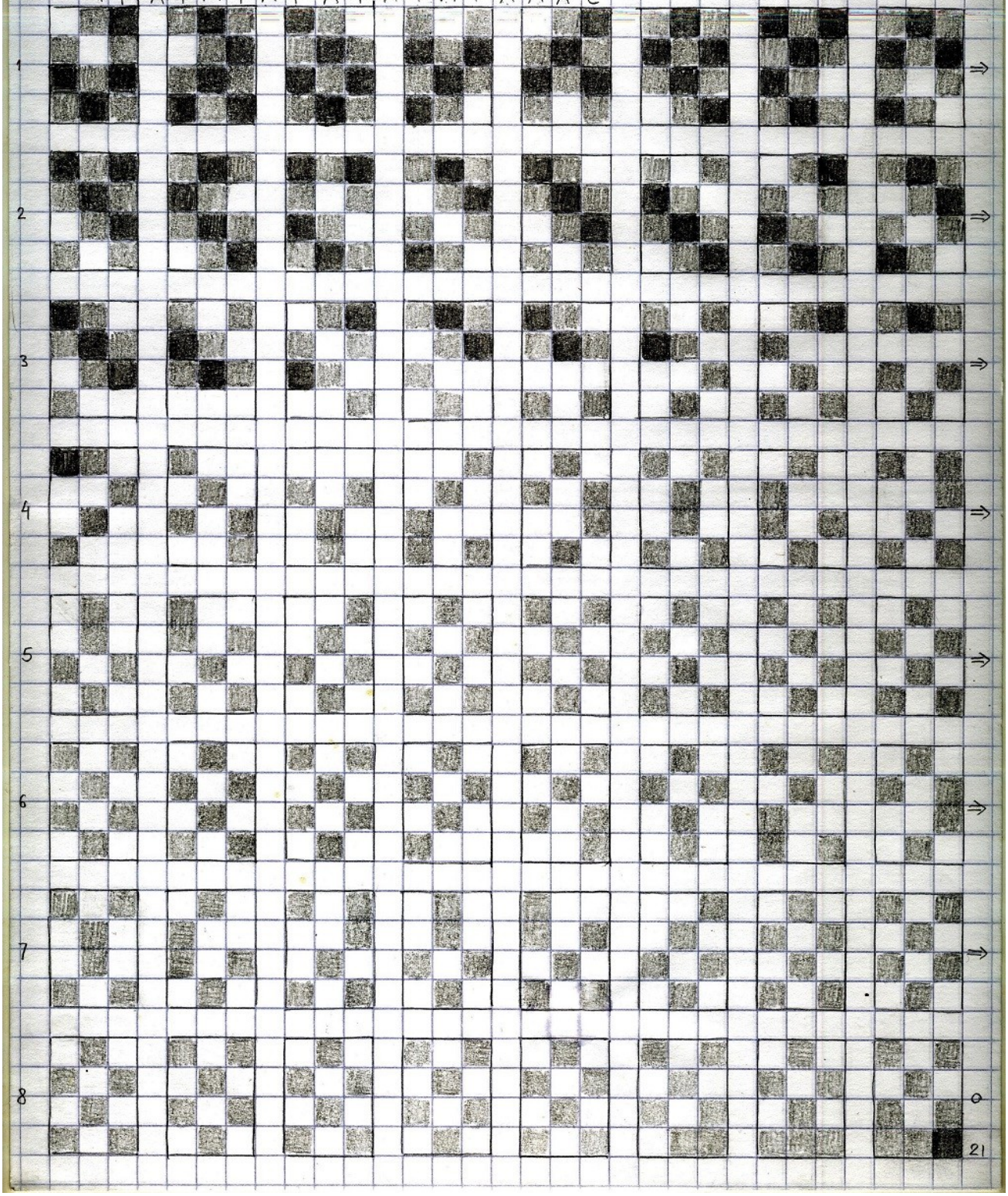
P.34466: A T A T A A A G T A A C T T T T T A A T A T T C A A
A C T T C C A C T A A T T T A G



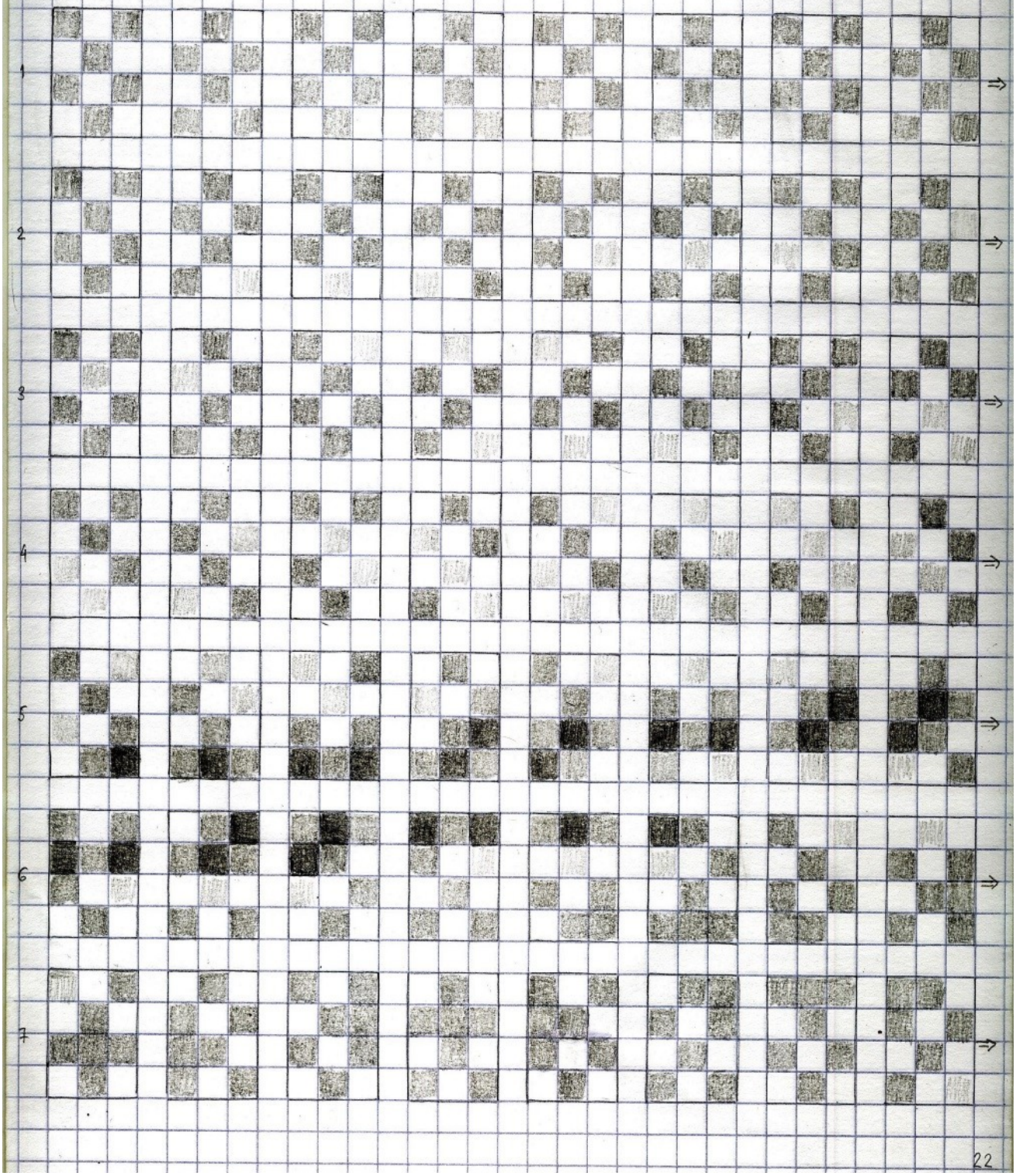
P.34603: T G T G T A T C T T T C T A T A T A T A A T T A T A T
T A A C T C C C

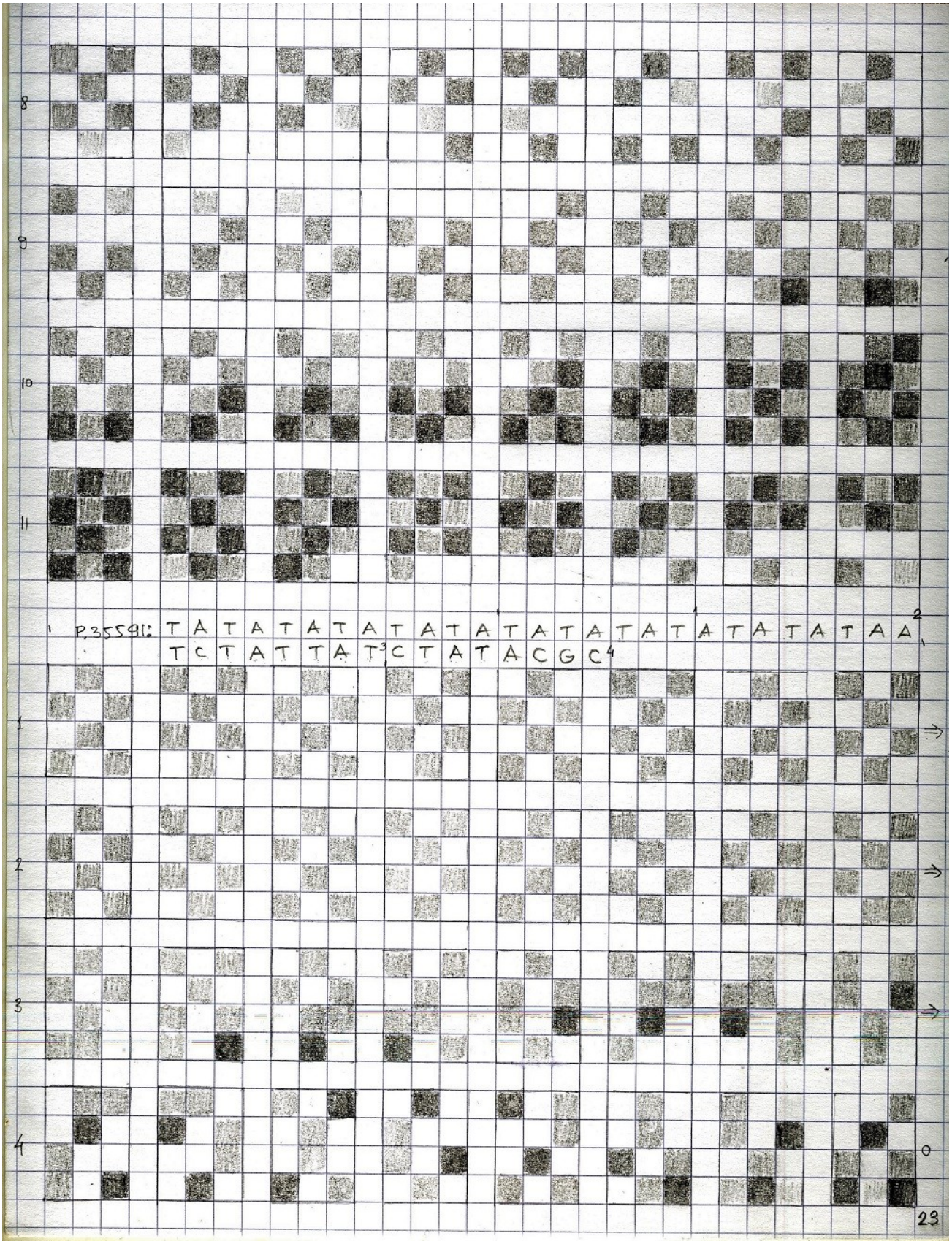


P.34994: TACATACACACACATACAT[1]ACATACAT[2]
TTATATAT[3]TATATATA[4]TATATATA[5]TATTAT
AT[6]ATATATATAT[7]ATATAAAC[8]

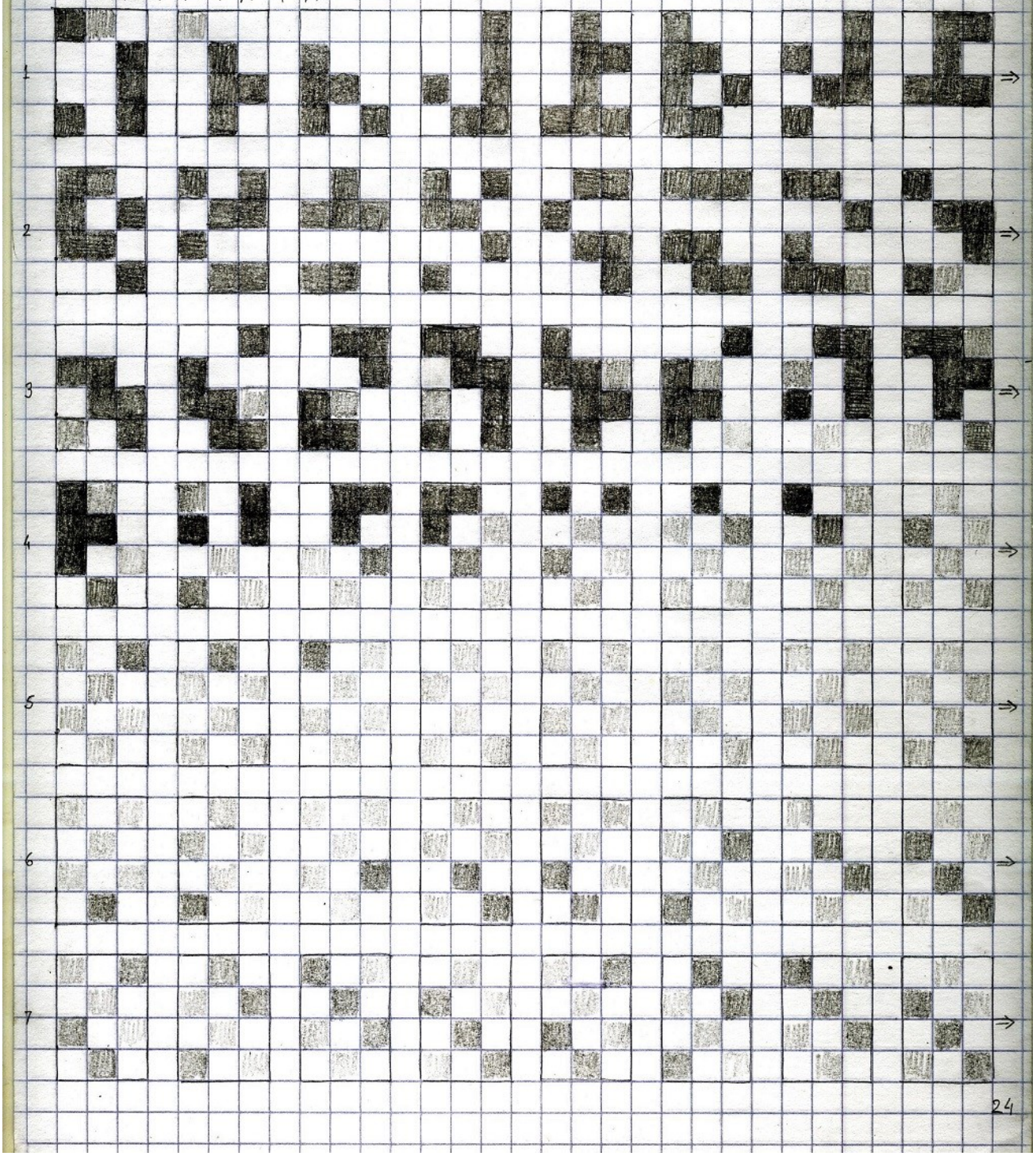


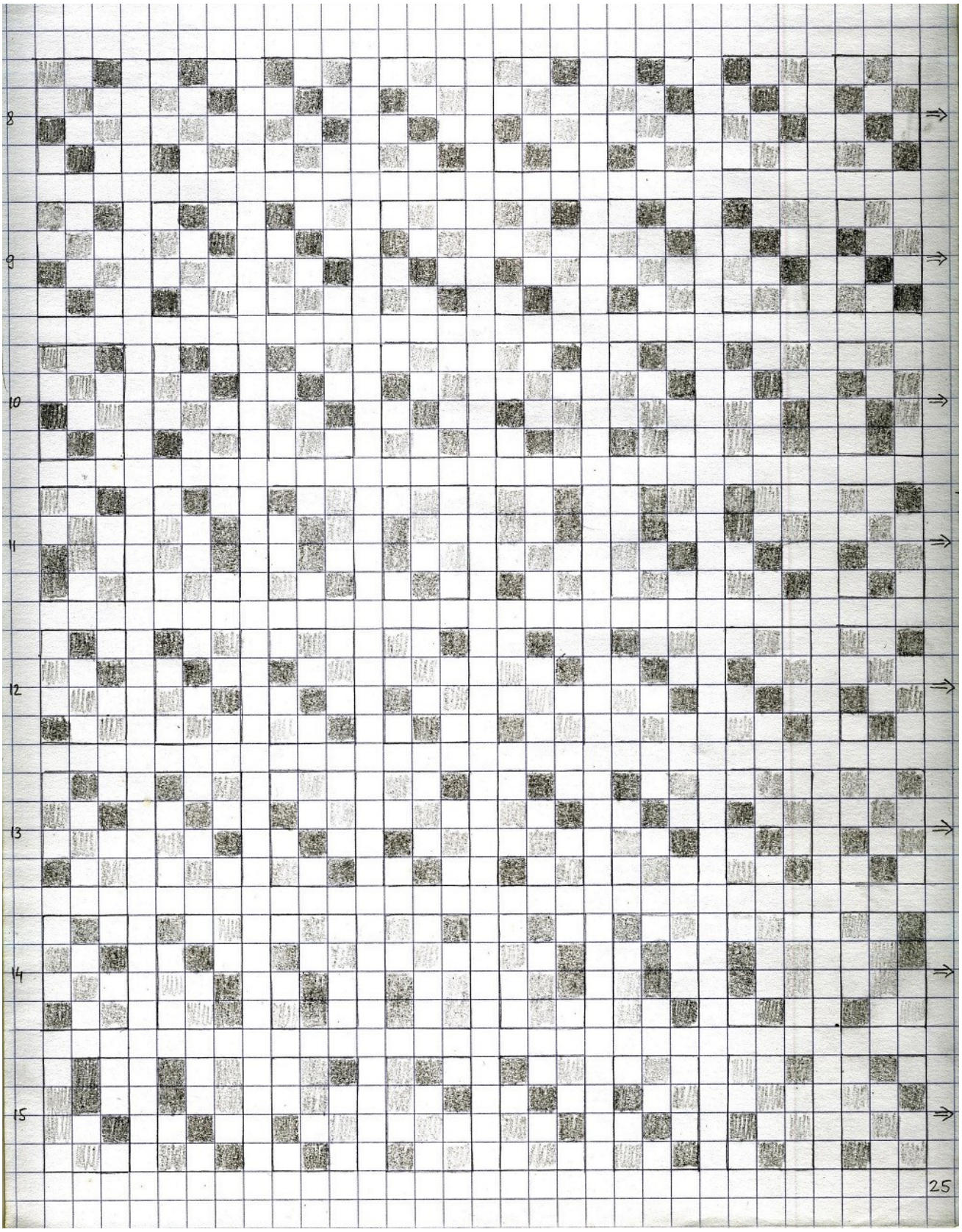
P.35488: ATATATATATATATATATATATGTATATA
TATGTATGTATGTATACACATGTATATAAA
TATATATATGTTTATATATATACACACA
CACACATTTATG

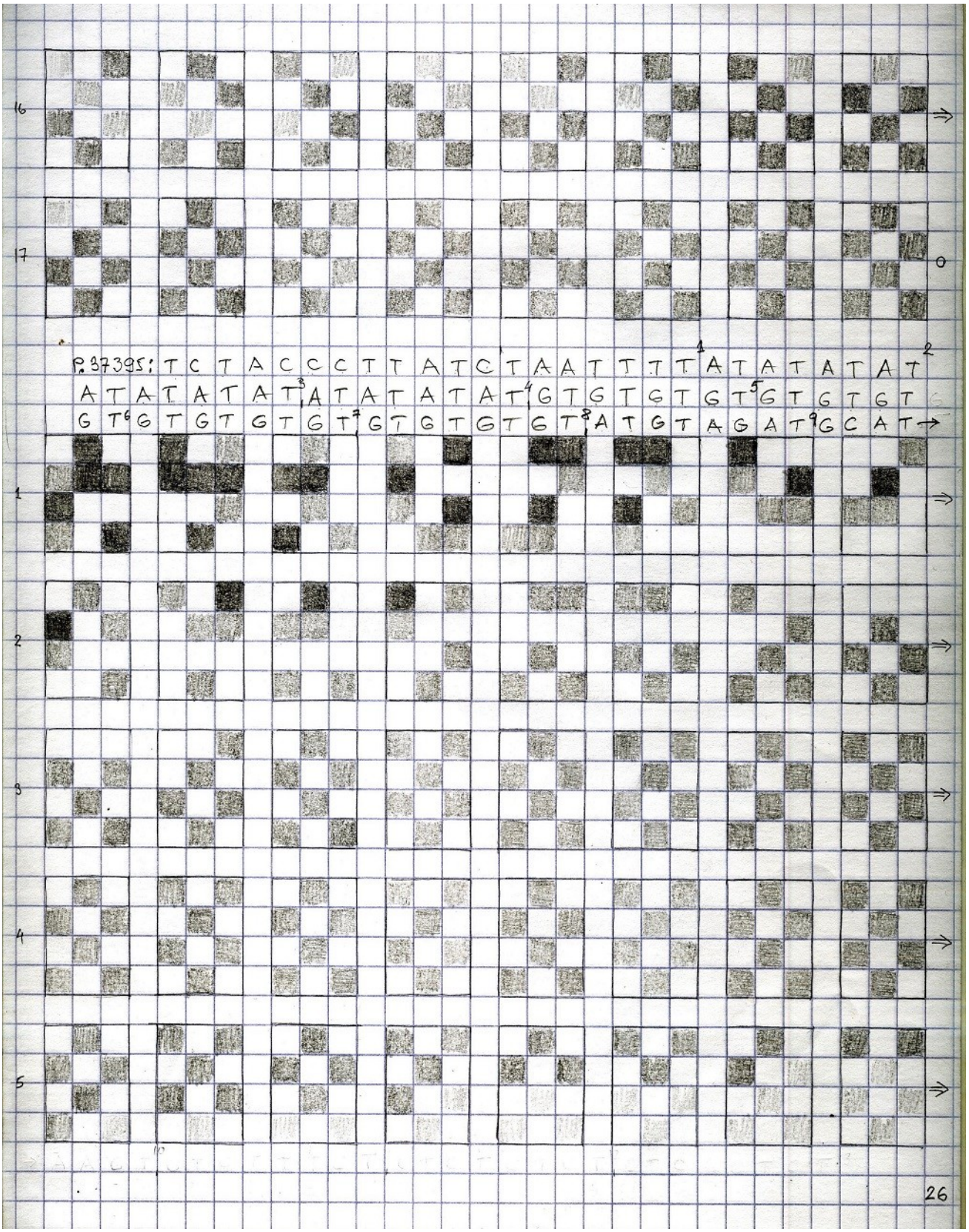


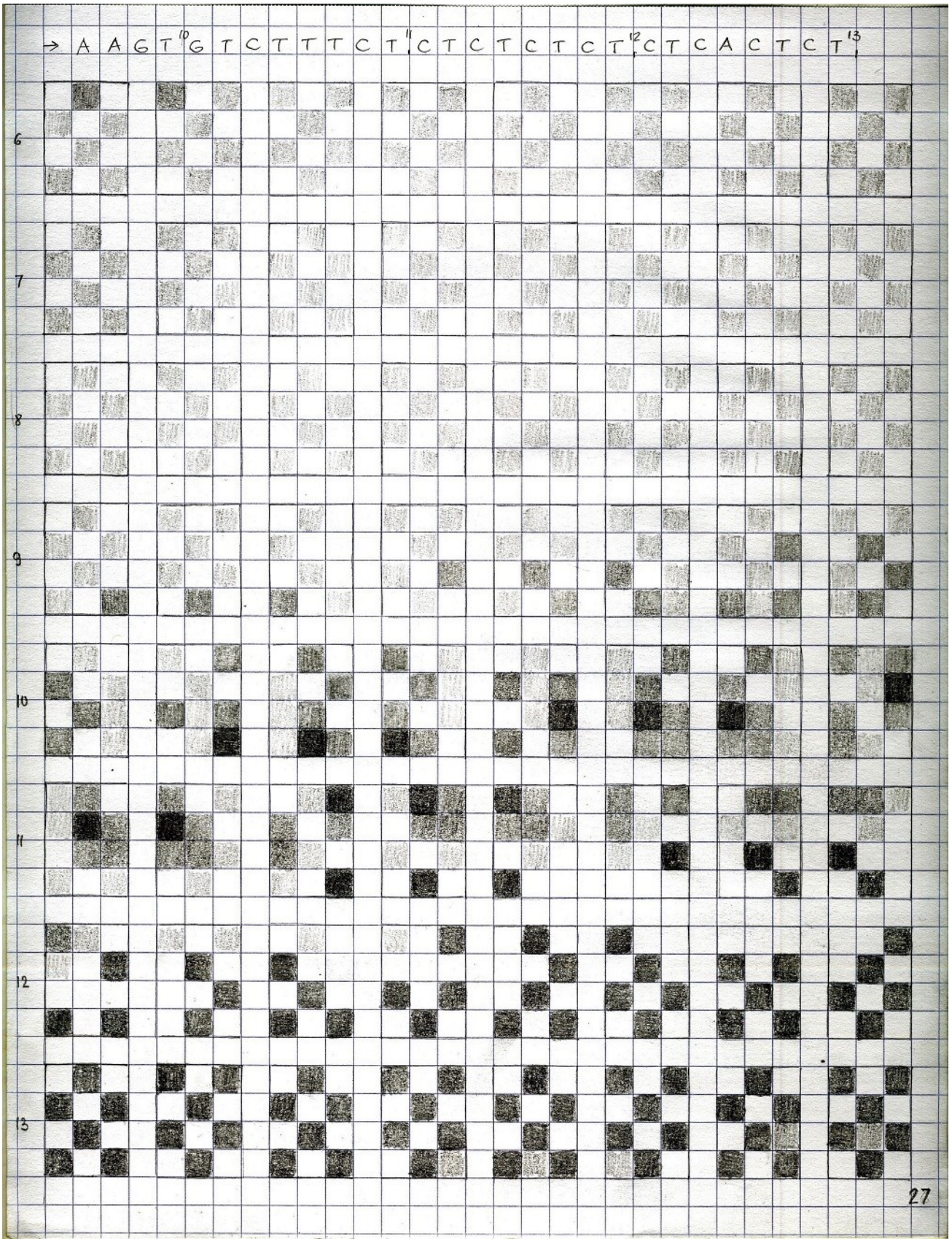

8
9
10
11
P3559: TATATATATATATATATATATATATAA
TCTATTATCTATACGC
1
2
3
4
23

P36103: CATTTCTTCCTCTCCCTTTCCTTCCAT
CCTCTGTATGTGTGTGTGTGTGTATGTATG
TATGTATGTATGTATGTATGTATGTATGTA
GTATGTATGTATGTATGTATGTATGTATGA
TGTATGTATGTATGTATATATATATATATA
TATATATA

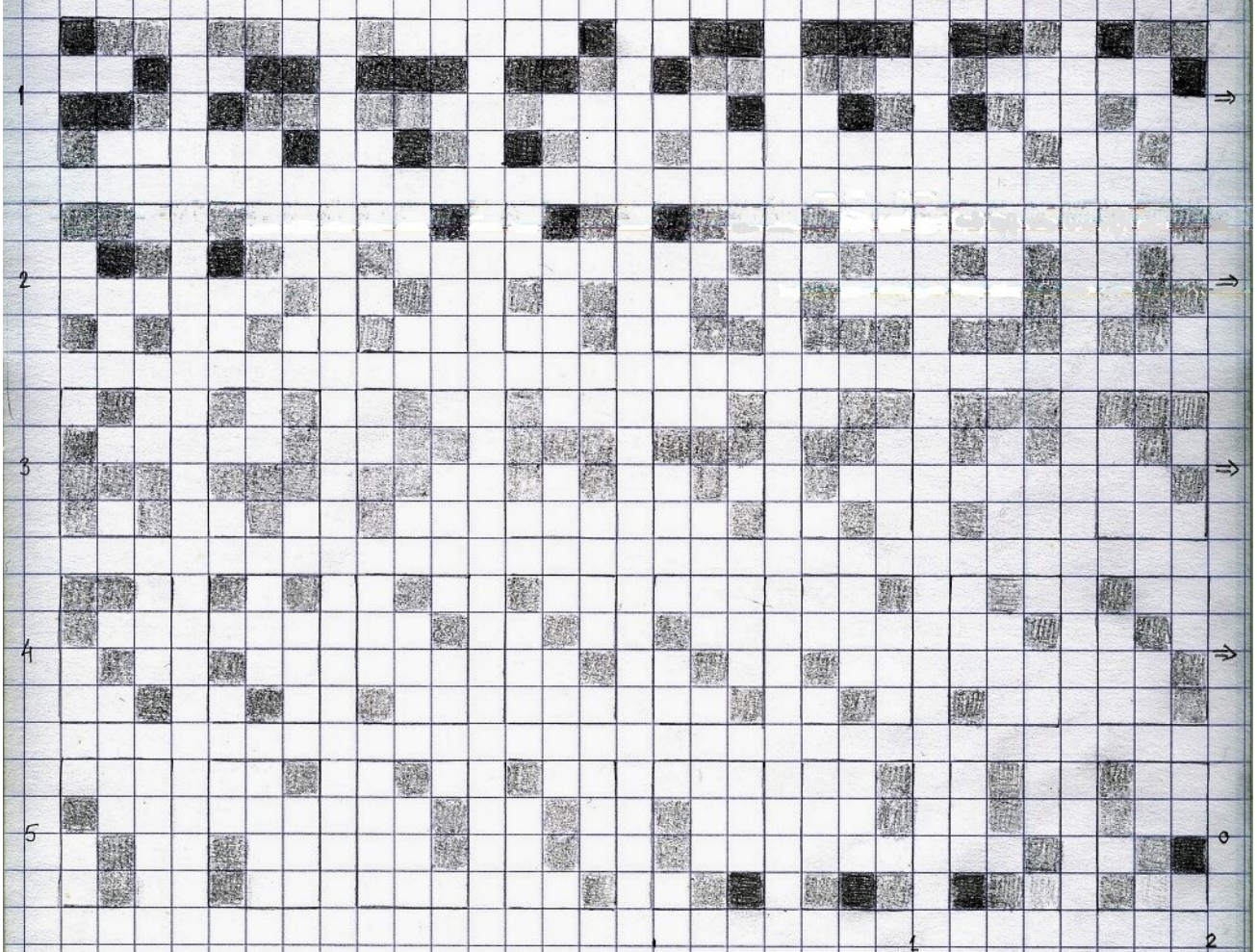



8
9
10
11
12
13
14
15

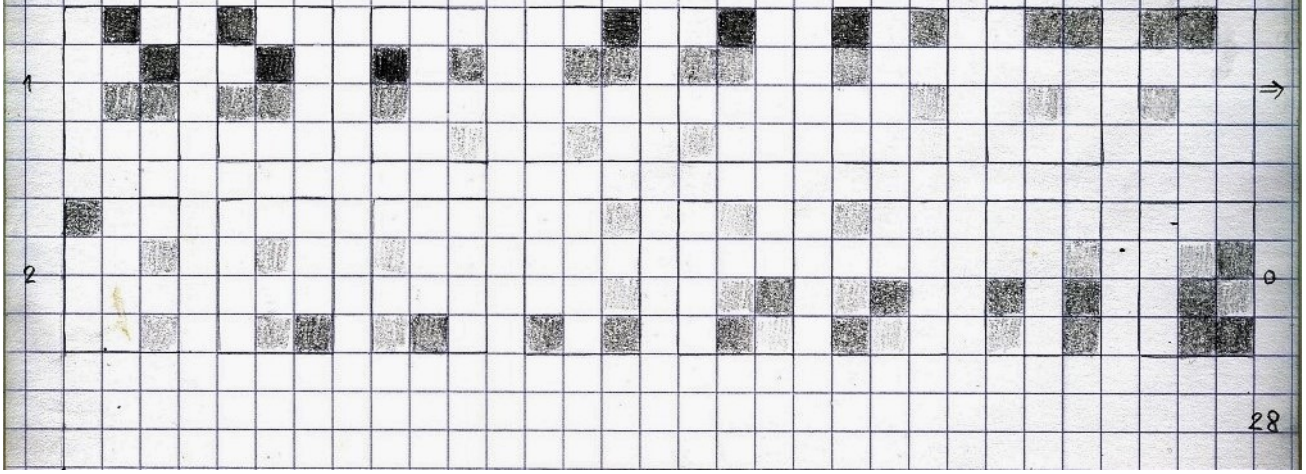

16
17
0
P.37395: TCTACCCTTATCTAATTTTATATATATAT
ATATATATATATATATATGTGTGTGTGTGTGTGT
GTGTGTGTGTGTGTGTGTGTGTGATGTAGATGCAT
1
2
3
4
5
26


→ A A G T G T C T T T C T C T C T C T C T C T C A C T C T
6
7
8
9
10
11
12
13

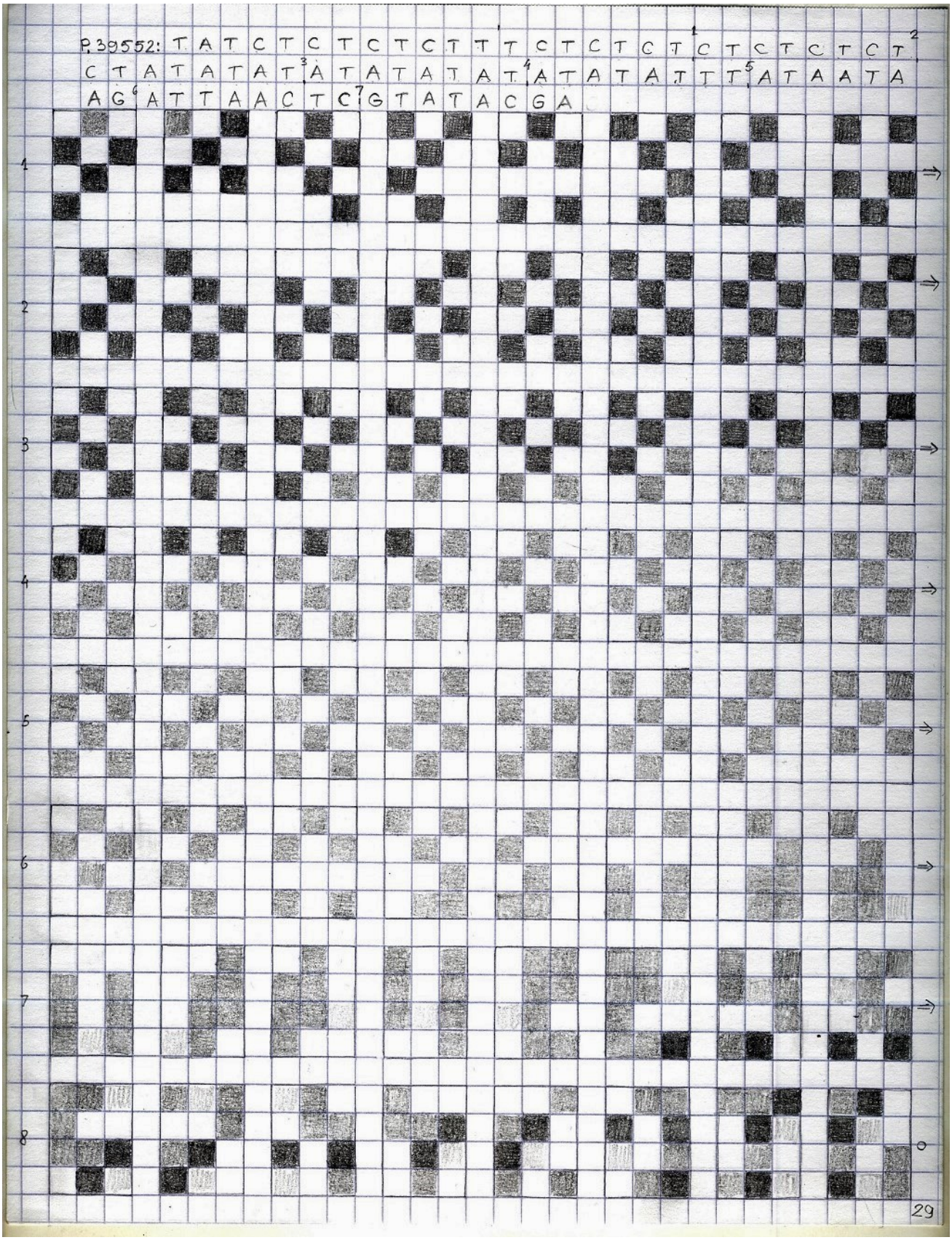
P.38148: C A A T T C C C A A T T C A T T T A T A T T A A A A T
A T T T A T T T A T T T A T T A T T T A C A G T

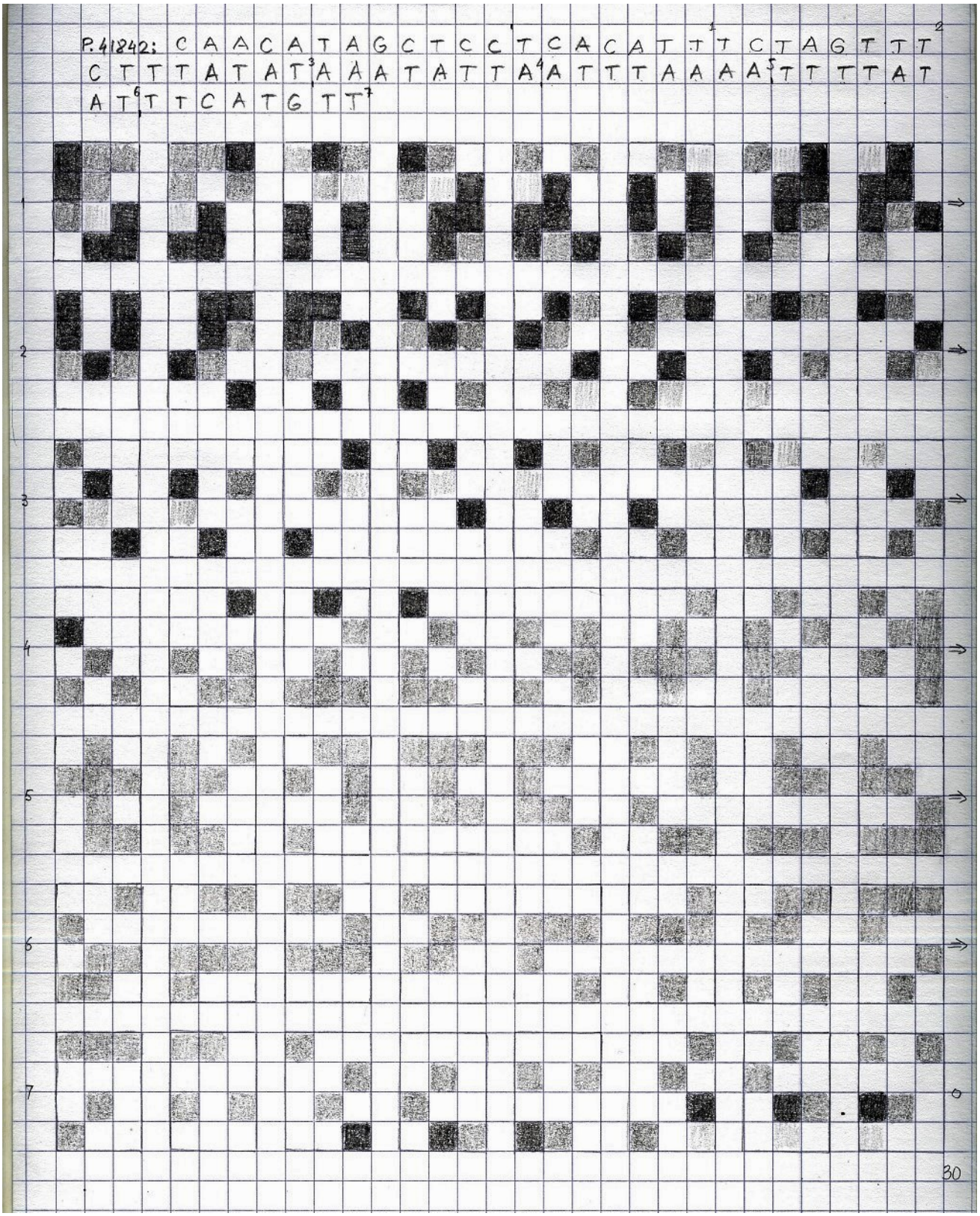


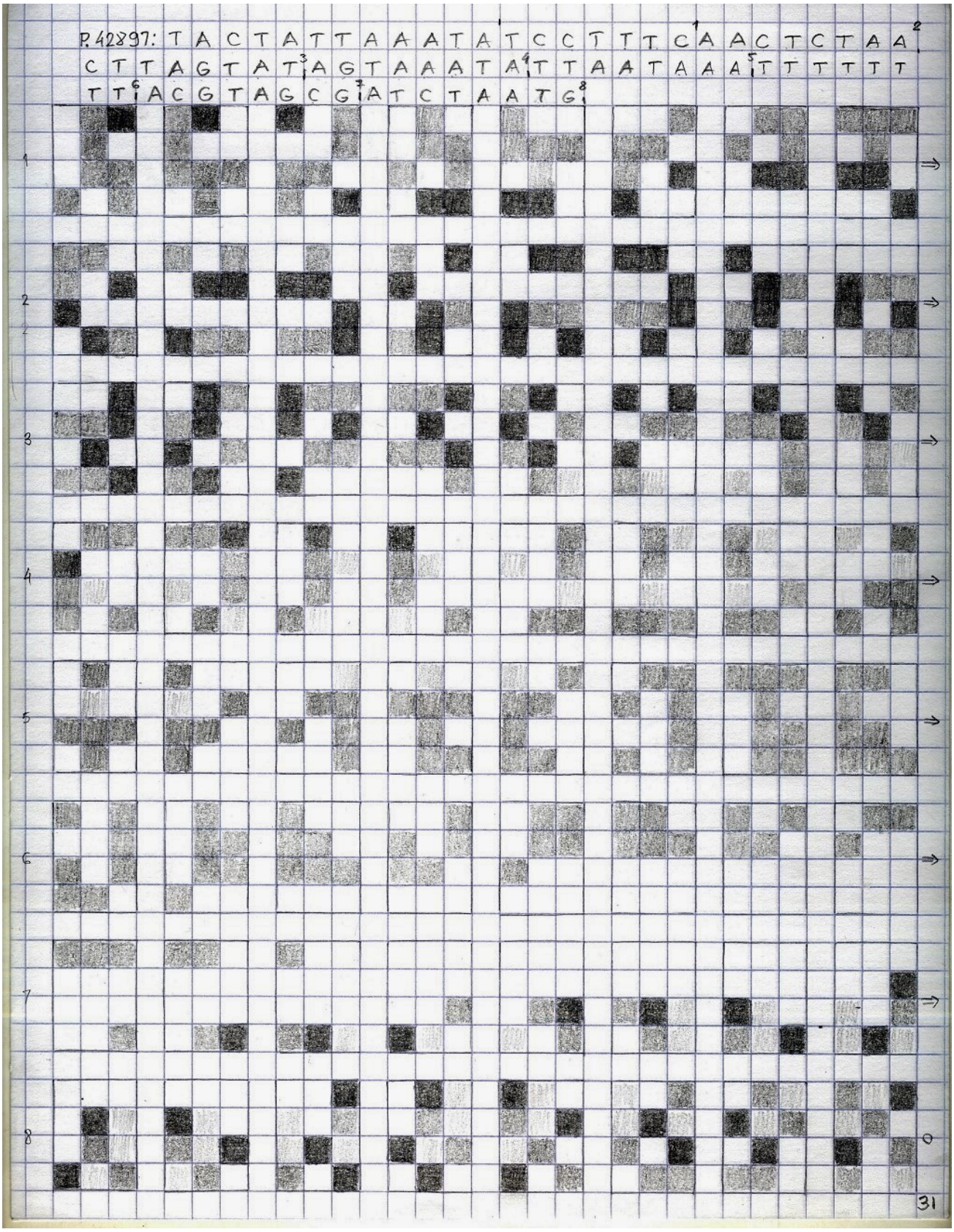
P.38335: T C T T T C T A A T T T T G T T T T T G A T A G T A A

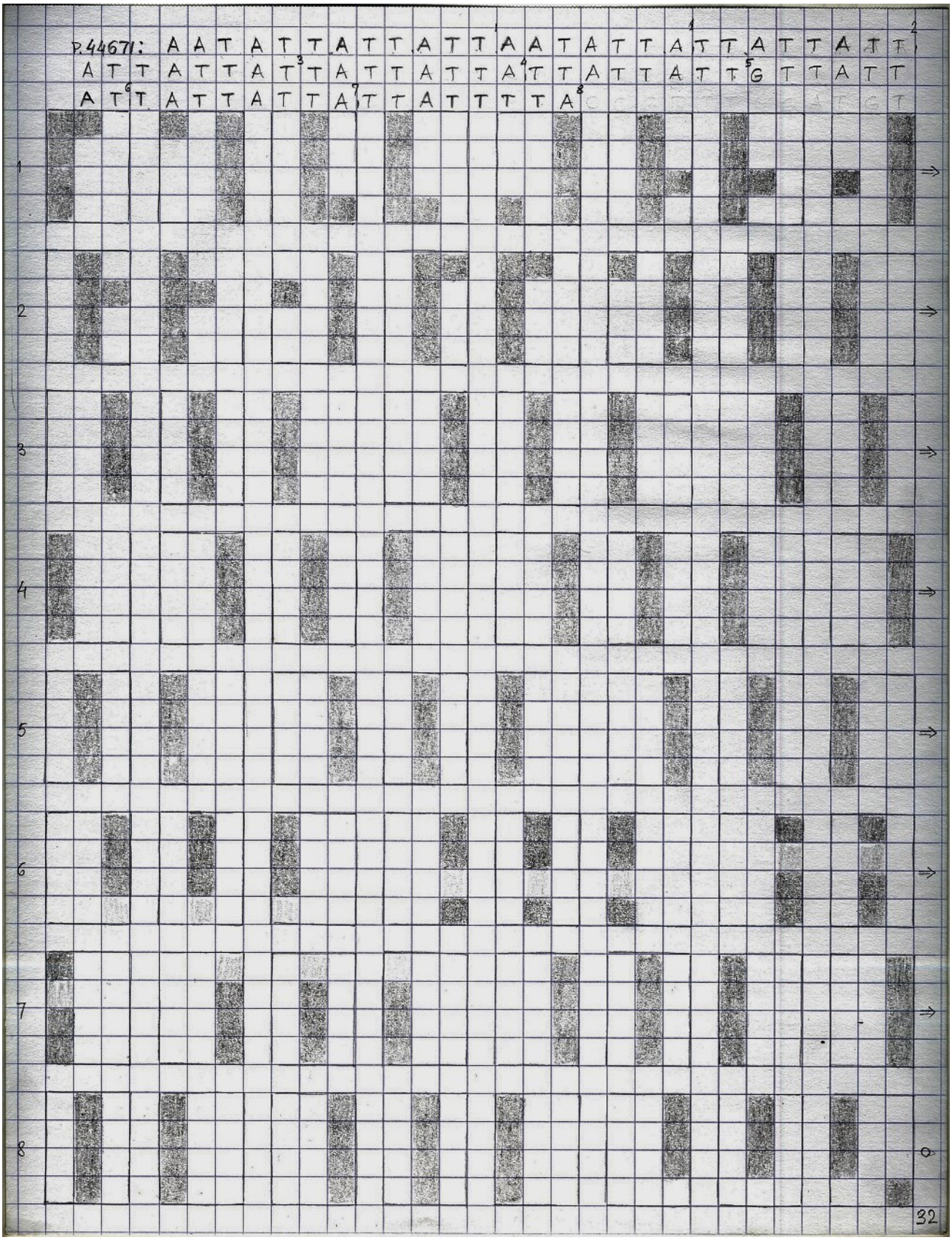


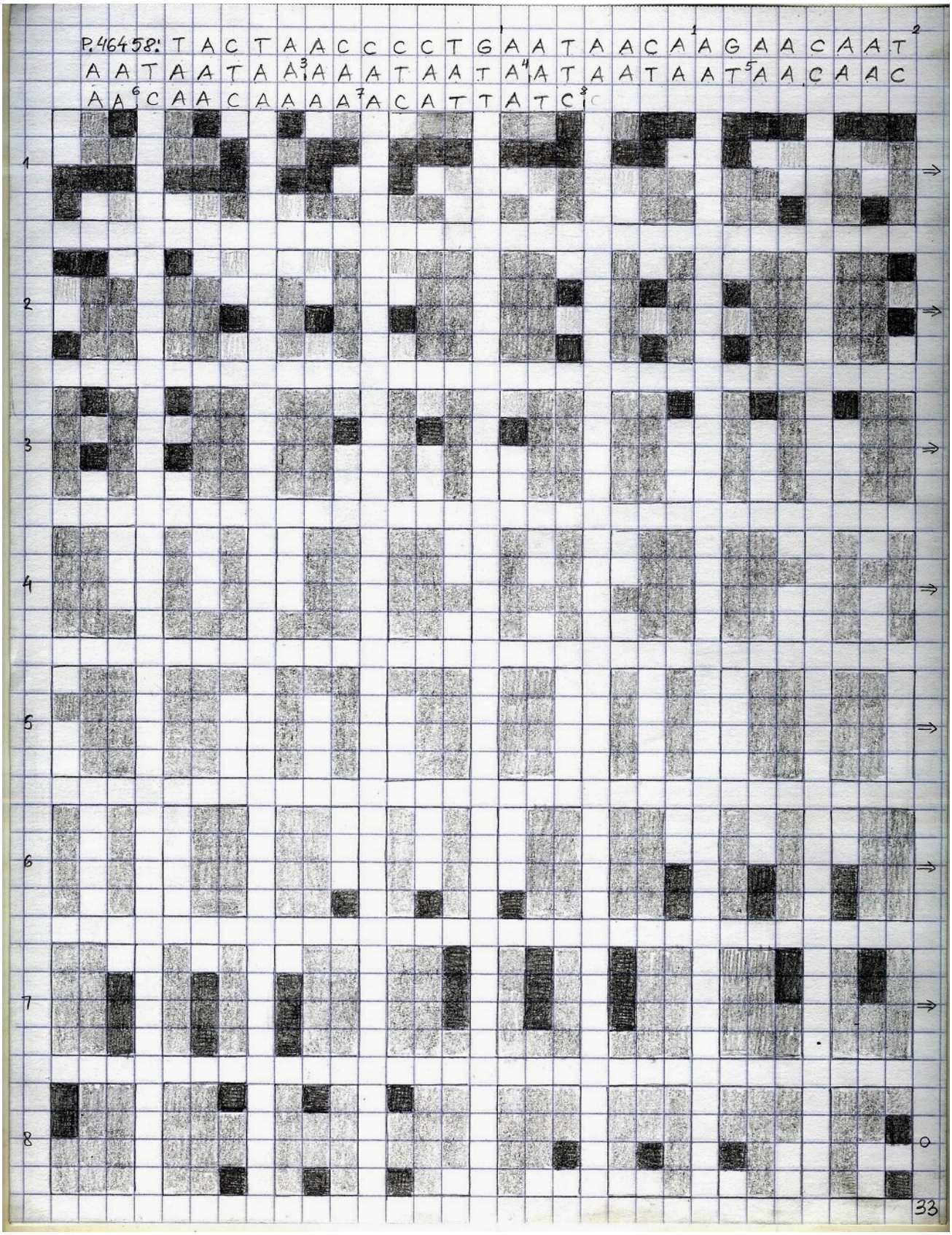
P.39552: TATCTCTCTCTTTCTCTCTCTCTCTCT
CTATATATATATATATATATTTATAATA
AGATTAACTCGTATACGA

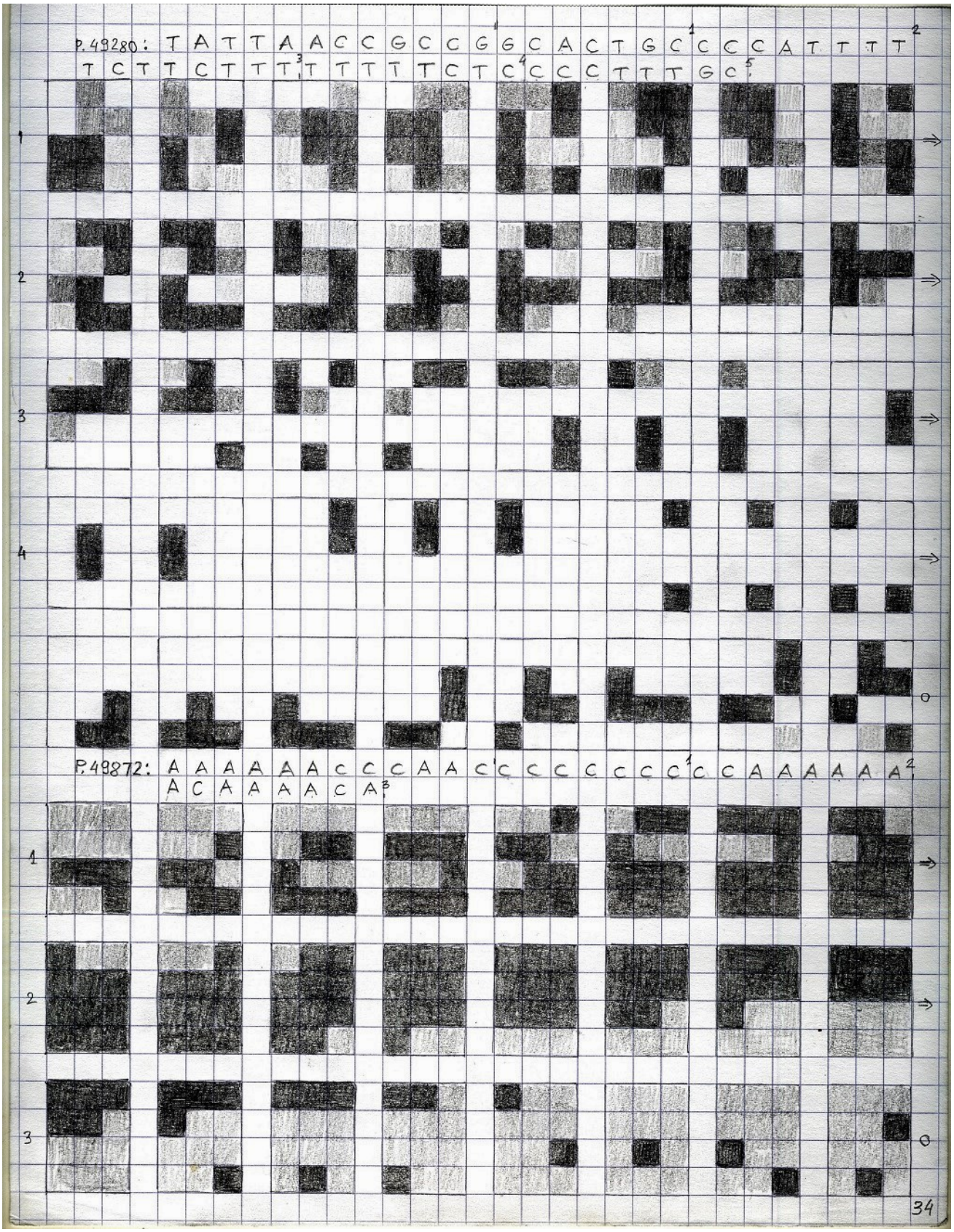


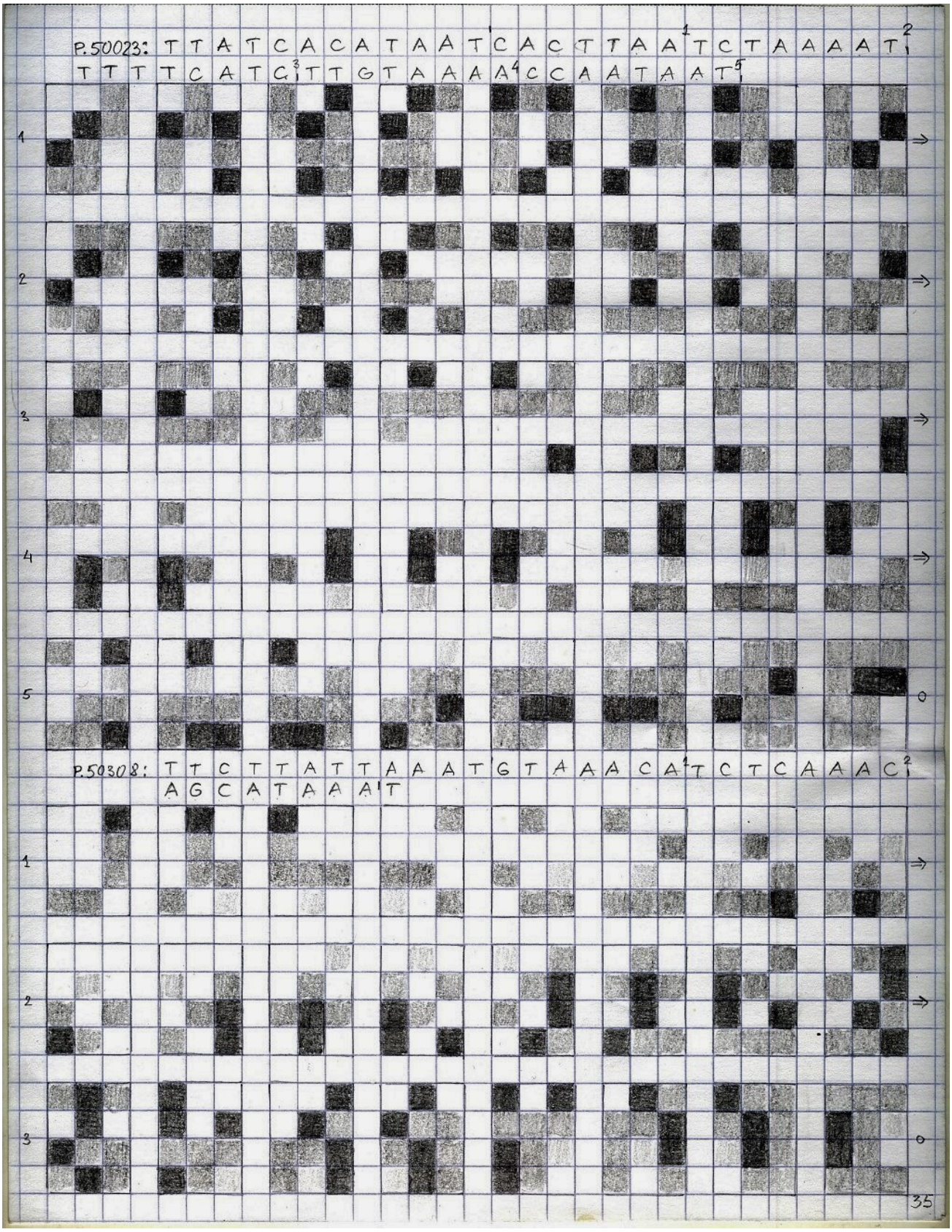

P.41842: C A A C A T A G C T C C T C A C A T T T C T A G T T T
C T T T A T A T A A A T A T T A A T T T A A A A T T T T A T
A T T T C A T G T T
1
2
3
4
5
6
7
1
2
3
4
5
6
7
30


P.42897: TACTATTAAATATCCTTTCAACTCTAA
CTTAGTATAGTAAATATTAATAAATTTTTTT
TTACGTAGCGATCTAATG
1
2
3
4
5
6
7
8
31

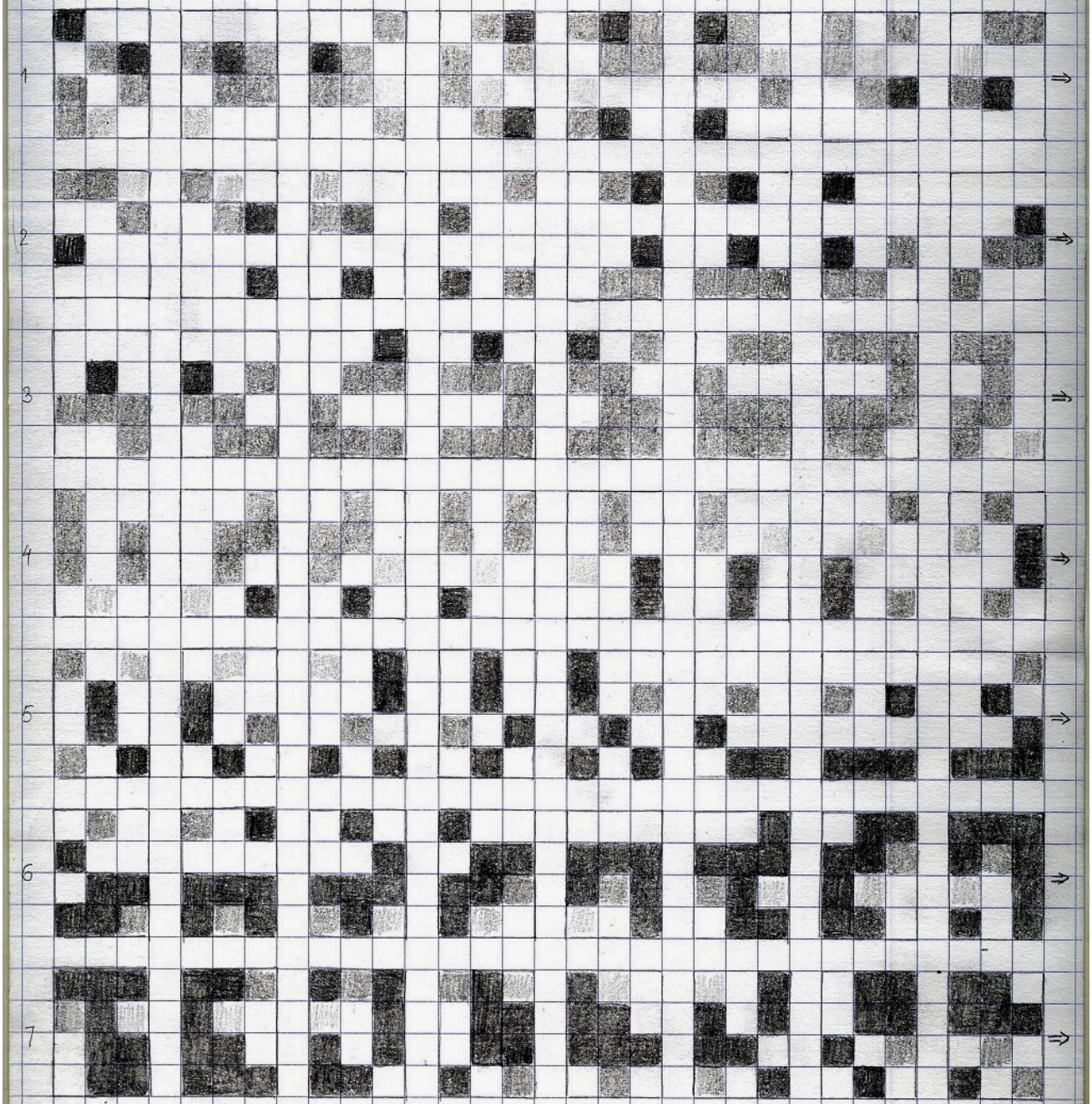

P.44671: AATATTATTATTAATATTATTATT

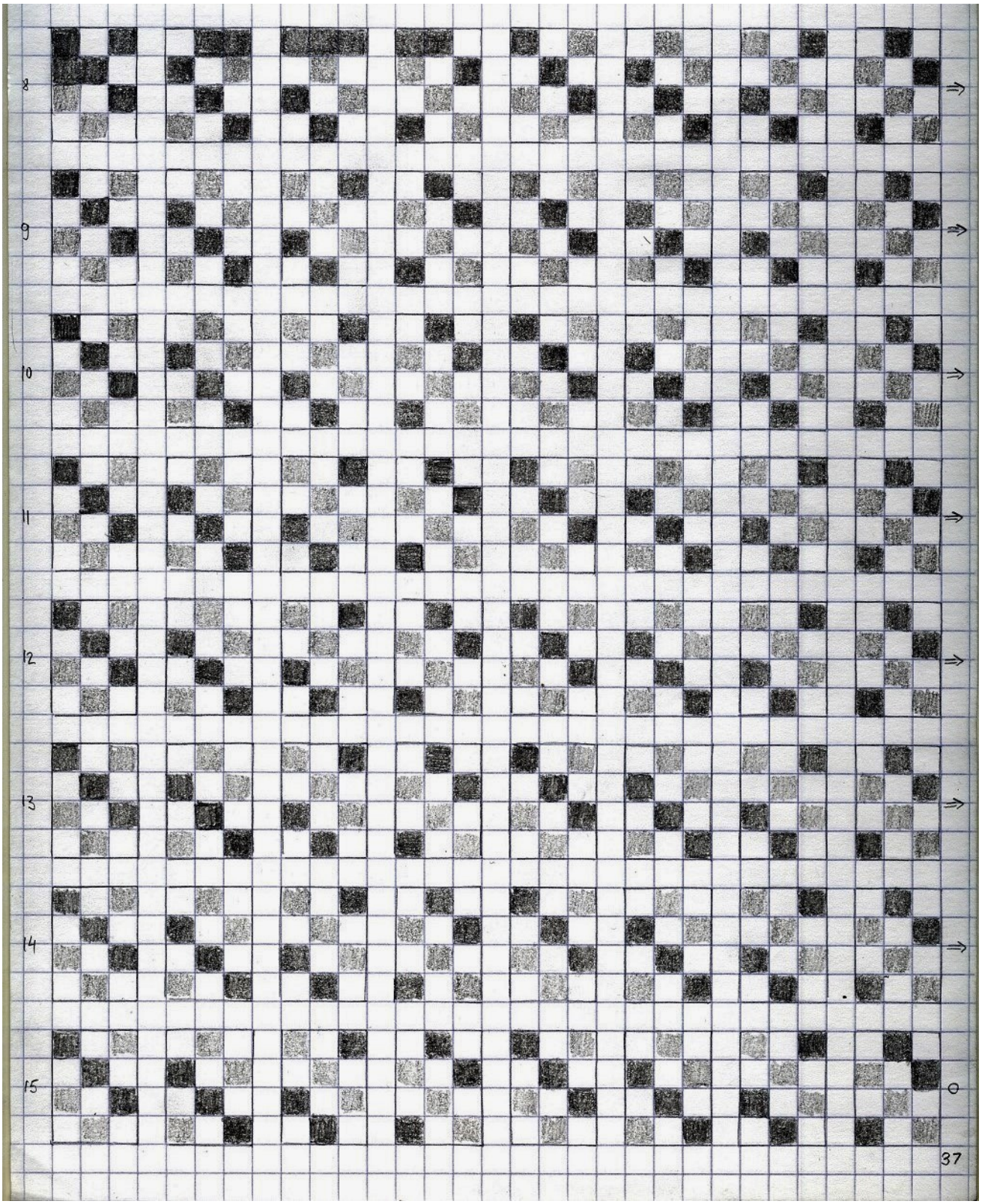

P.46458: TACTAACCCCTGAATAACAAGAACAAT
AATAATAAAAAATAATAATAATAATAACAAC
AA CAACAAAAACATTATC
1
2
3
4
5
6
7
8

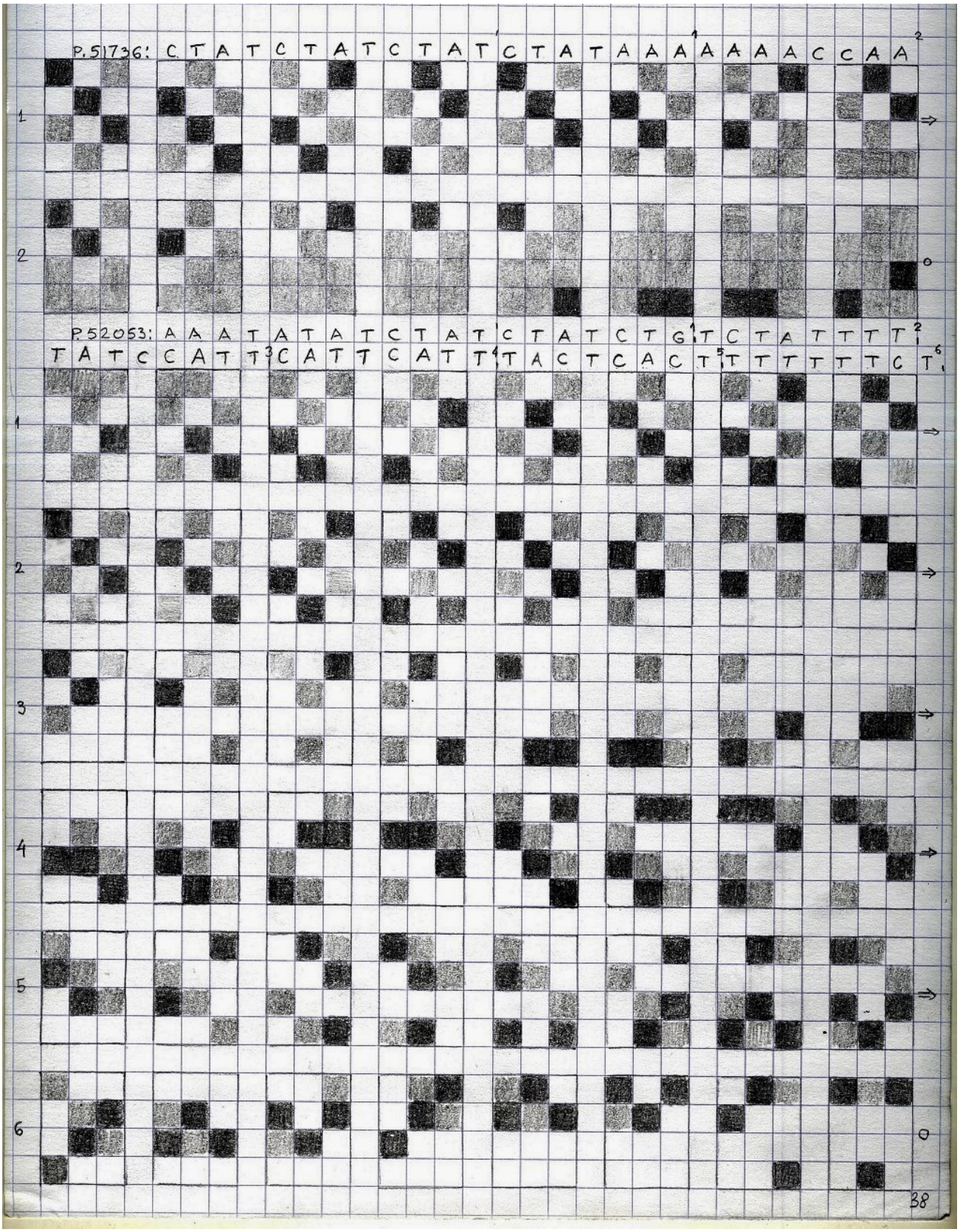

P.49280: TATTAACCGCCGGCACTGCCCATTTT
TCTTCTTTTTTTTTCTCCCCTTTGC
1
2
3
4
P.49872: AAAAAACCCAACCCCCCCCCCAAAAAAAA
ACAAAAACA
1
2
3

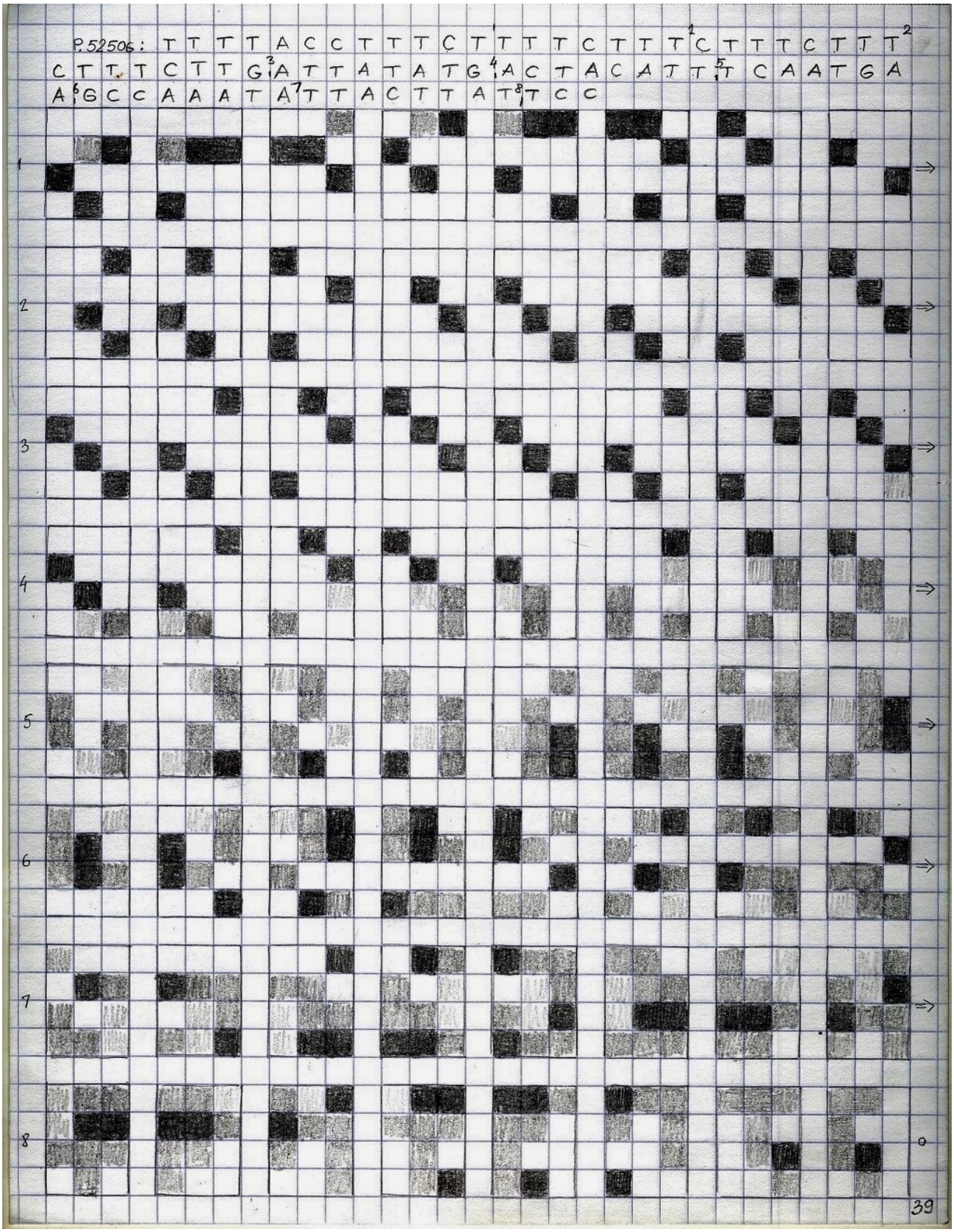

P.50023: T T A T C A C A T A A T C A C T T A A T C T A A A A T
T T T T C A T C T T G T A A A A C C A A T A A T
1
2
3
4
5
P.50308: T T C T T A T T A A A T G T A A A C A T C T C A A A C
A G C A T A A A T
1
2
3
35

P.51620: CTTTACATAAGTTACTTTTCTAAATT
AAAAAAATGTCTTCTATCTCTCCCCACGTCC
TCCCTATCTATCTATCTATCTATCTATCTA
TCTATCTATCTATCTATCTATCTATCTATC
TATCTATCTATCTA

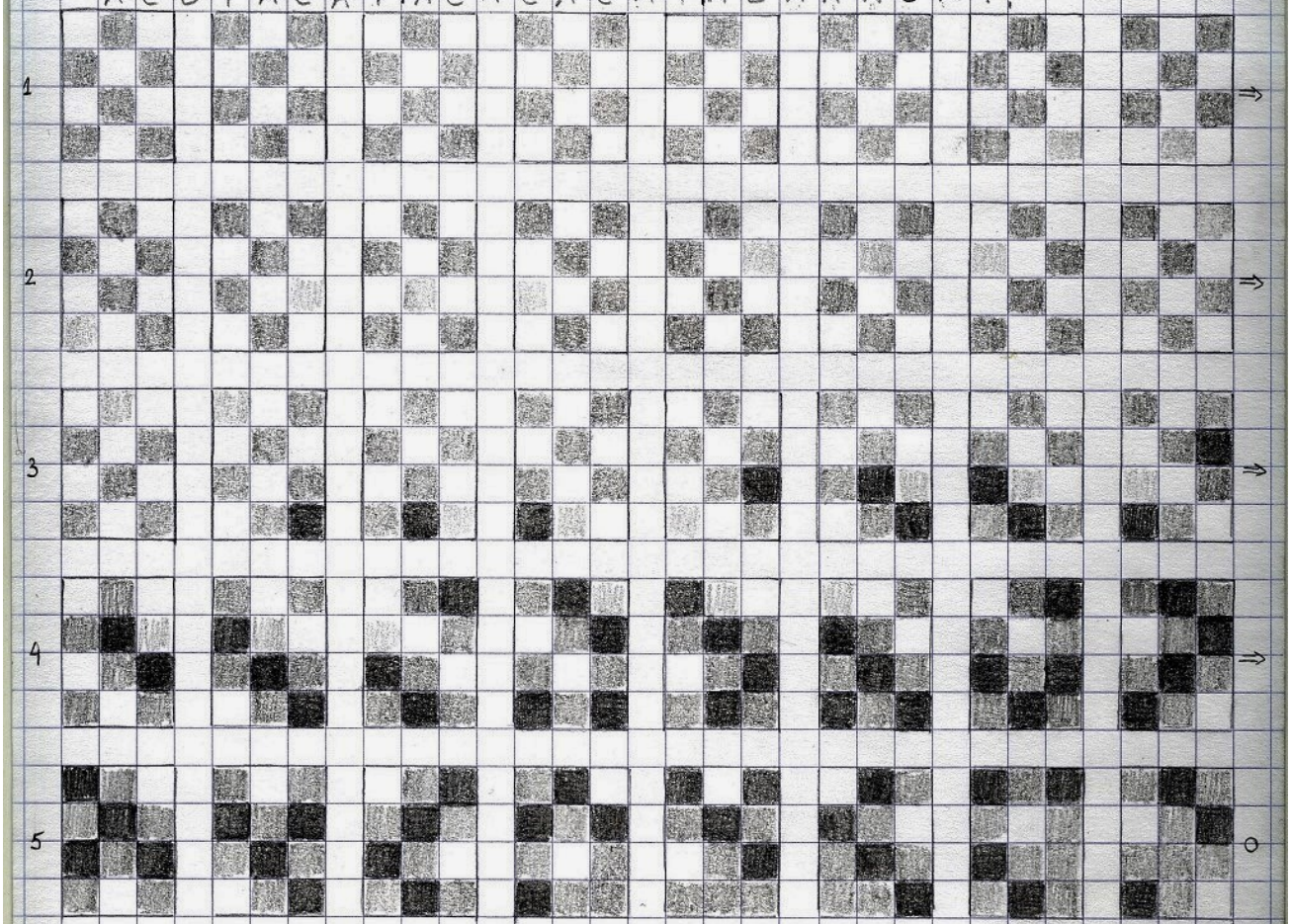



8
9
10
11
12
13
14
15

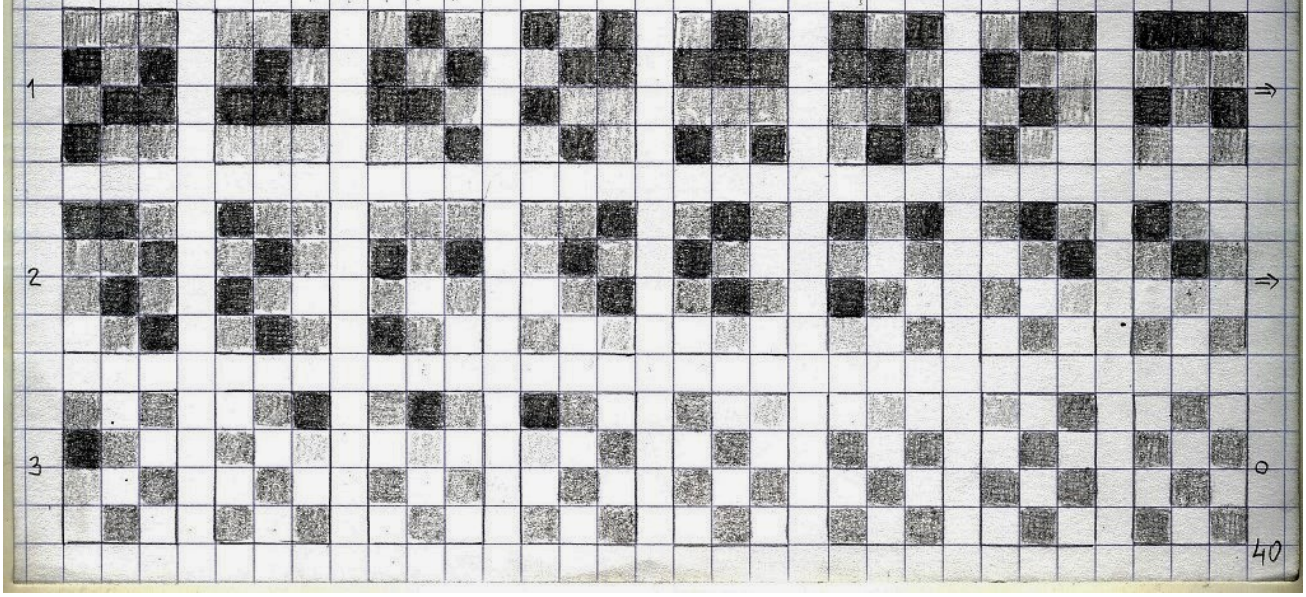

P.51736: C T A T C T A T C T A T C T A T A A A A A A A C C A A
1
2
P.52053: A A A T A T A T C T A T C T A T C T G T C T A T T T T
T A T C C A T T C A T T C A T T T A C T C A C T T T T T T C T
1
2
3
4
5
6
38

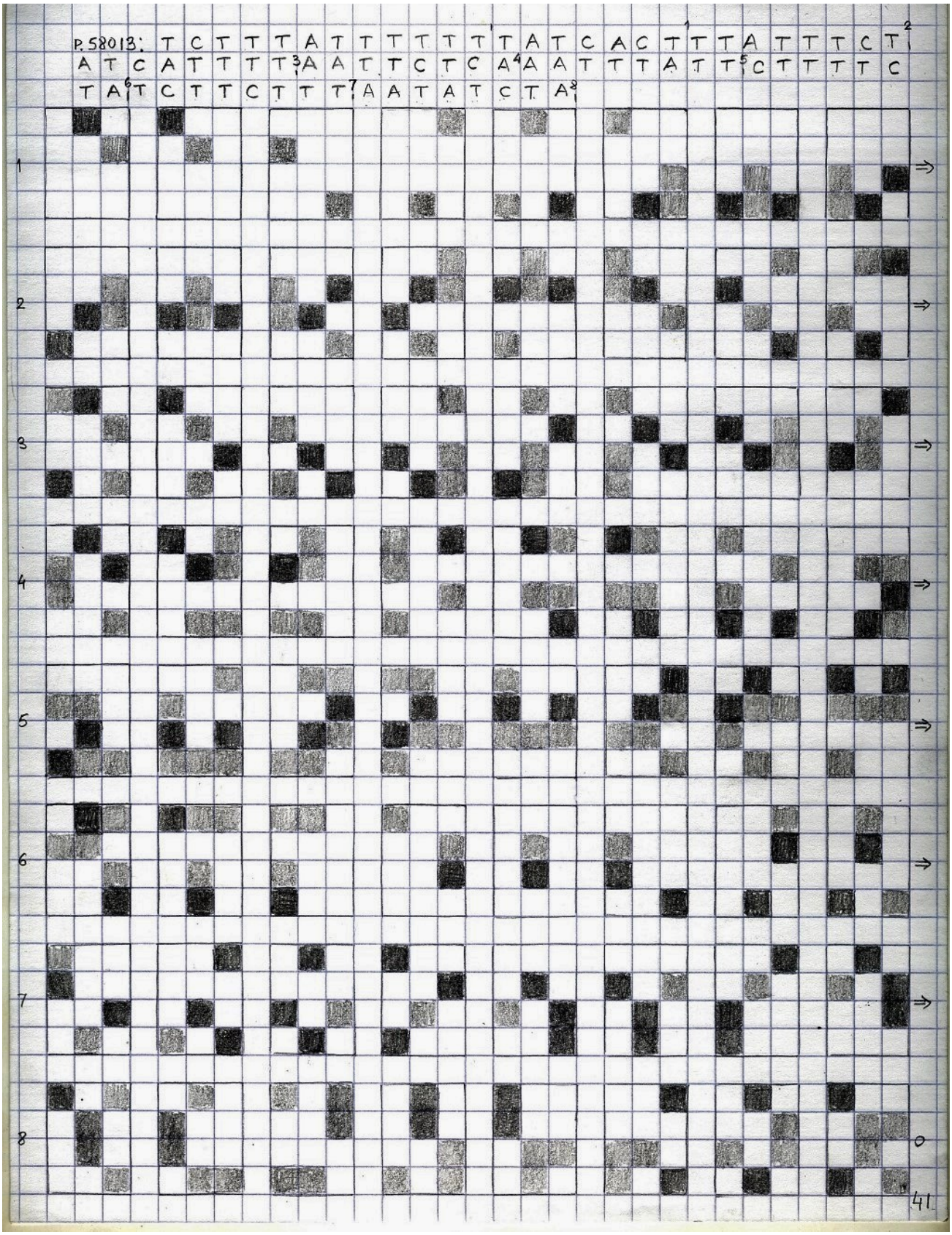

P.52506: T T T T A C C T T T C T T T T C T T T C T T T C T T T
C T T T C T T G A T T A T A T G A C T A C A T T T C A A T G A
A G C C A A A T A T T A C T T A T T C C
1
2
3
4
5
6
7
8
1
2
3
4
5
6
7
8
o
39

P.53876: T A T A T A T A T A T A T A T A T G T A T A T A T A T
A C G T A C A T A C A C A C A T A C A A A C A T

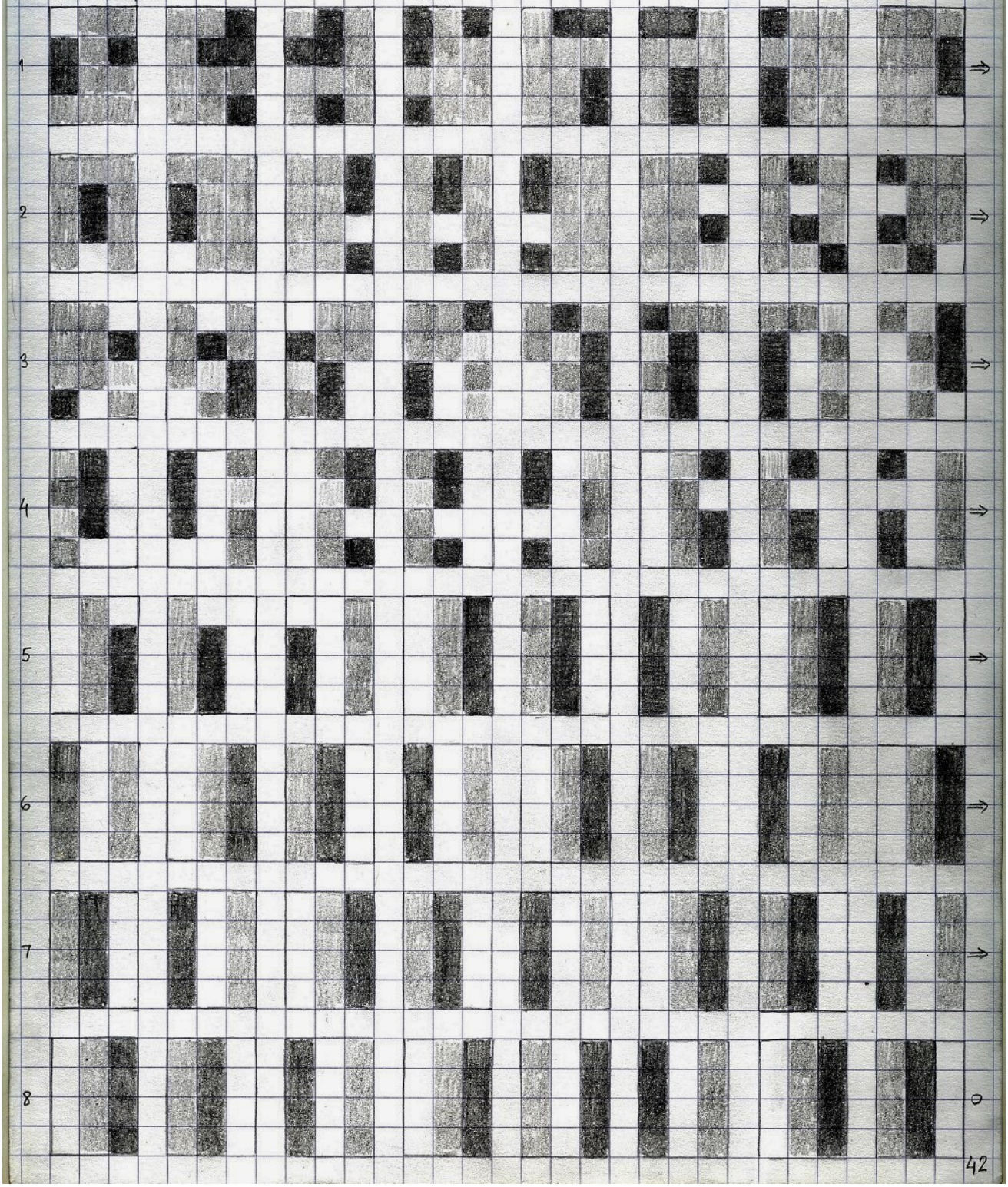


P.53780: A A A C A C A C C C A A A C A C A T A C A T G T A T A
T A T A T A T A

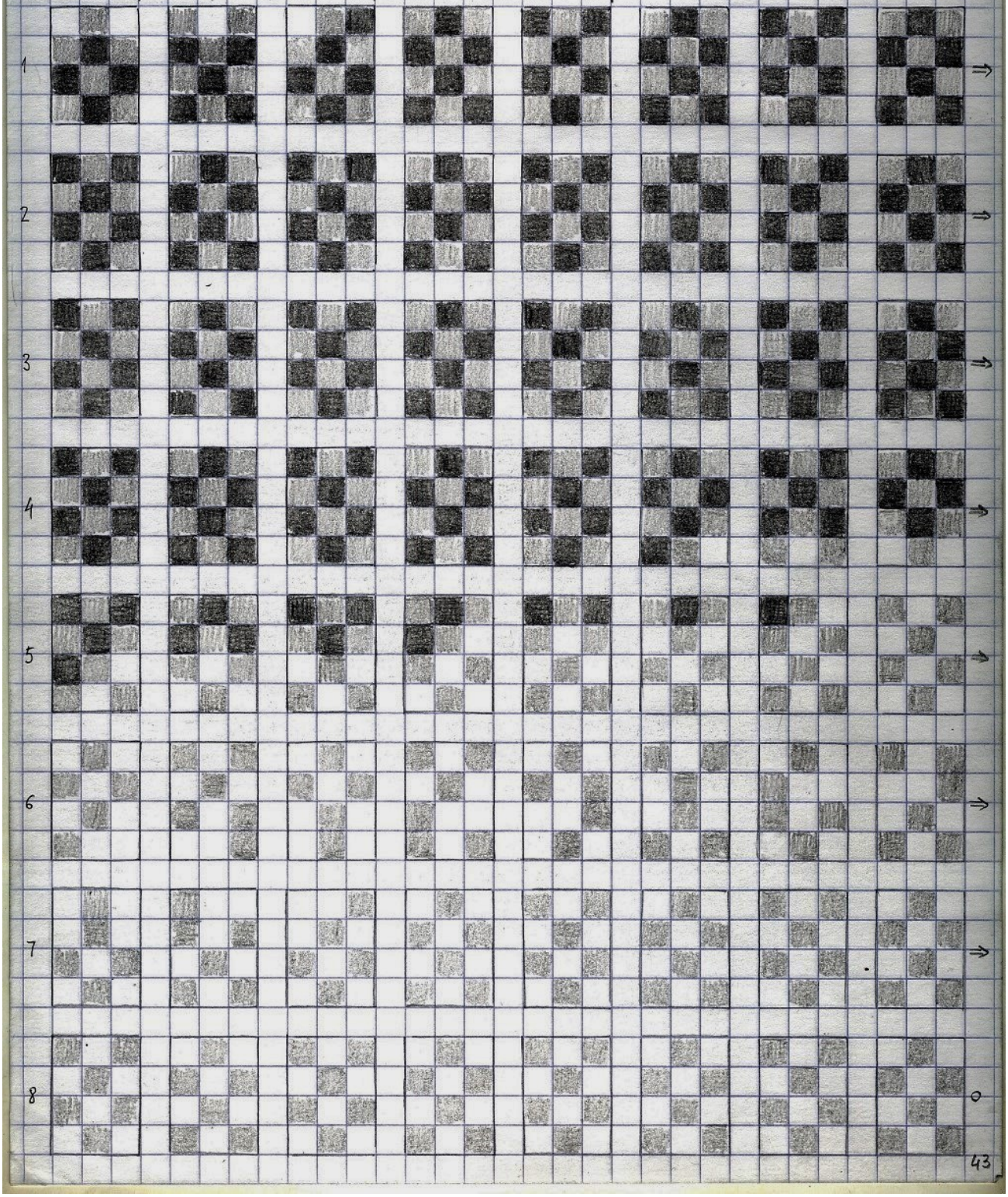


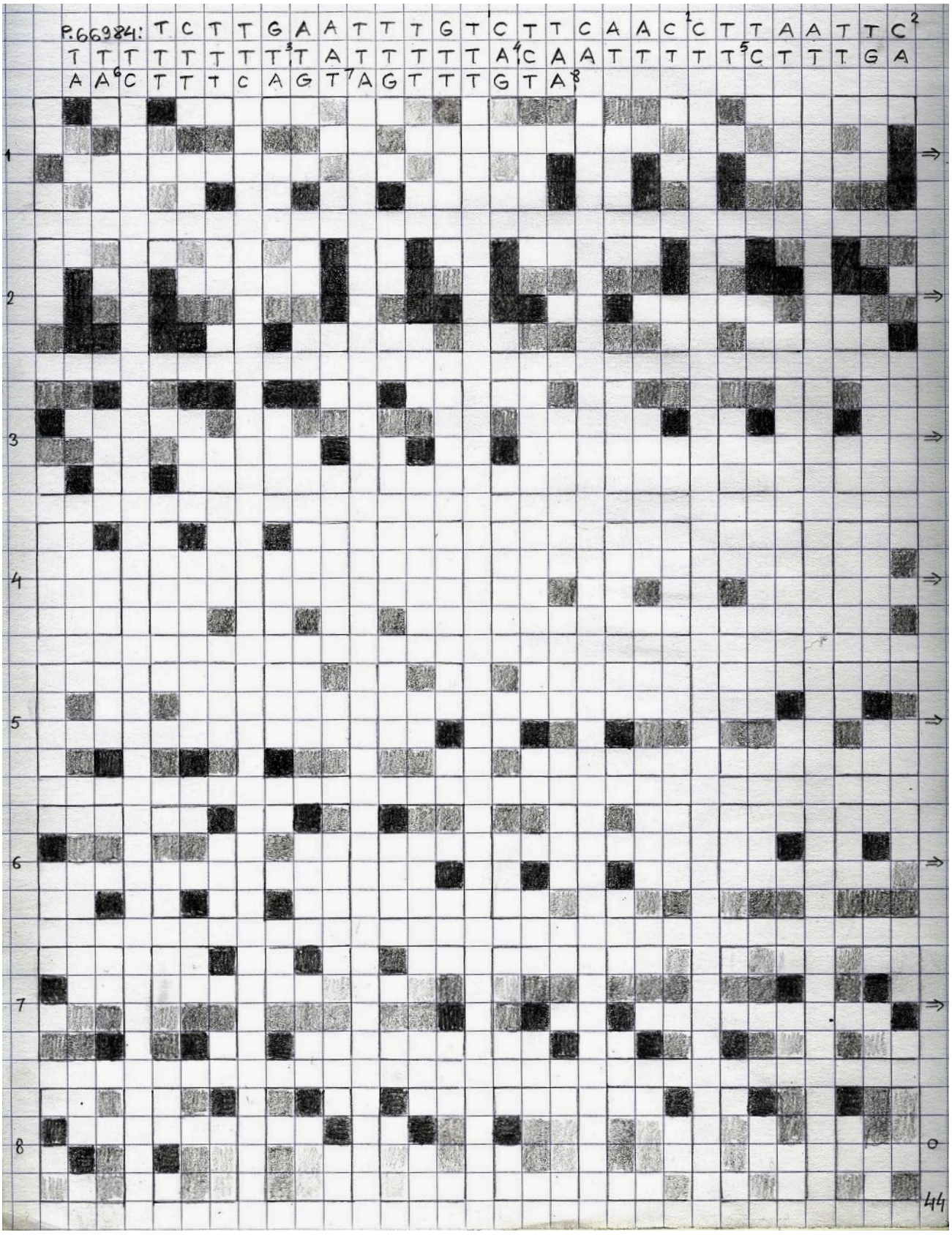

P.58013: TCTTTATTTTTTTATCACTTTATTTCT
ATCATTTTAATTCTCAAATTTATTCTTTTC
TATCTTTCTTTTAATATCTA
1
2
3
4
5
6
7
8
41

P. 61628: TAACACCAAAAACAACAATAACAAGCT
ACTGCTATTACTACTACTACTACTACT
ACTACTACTACTACTACTACT

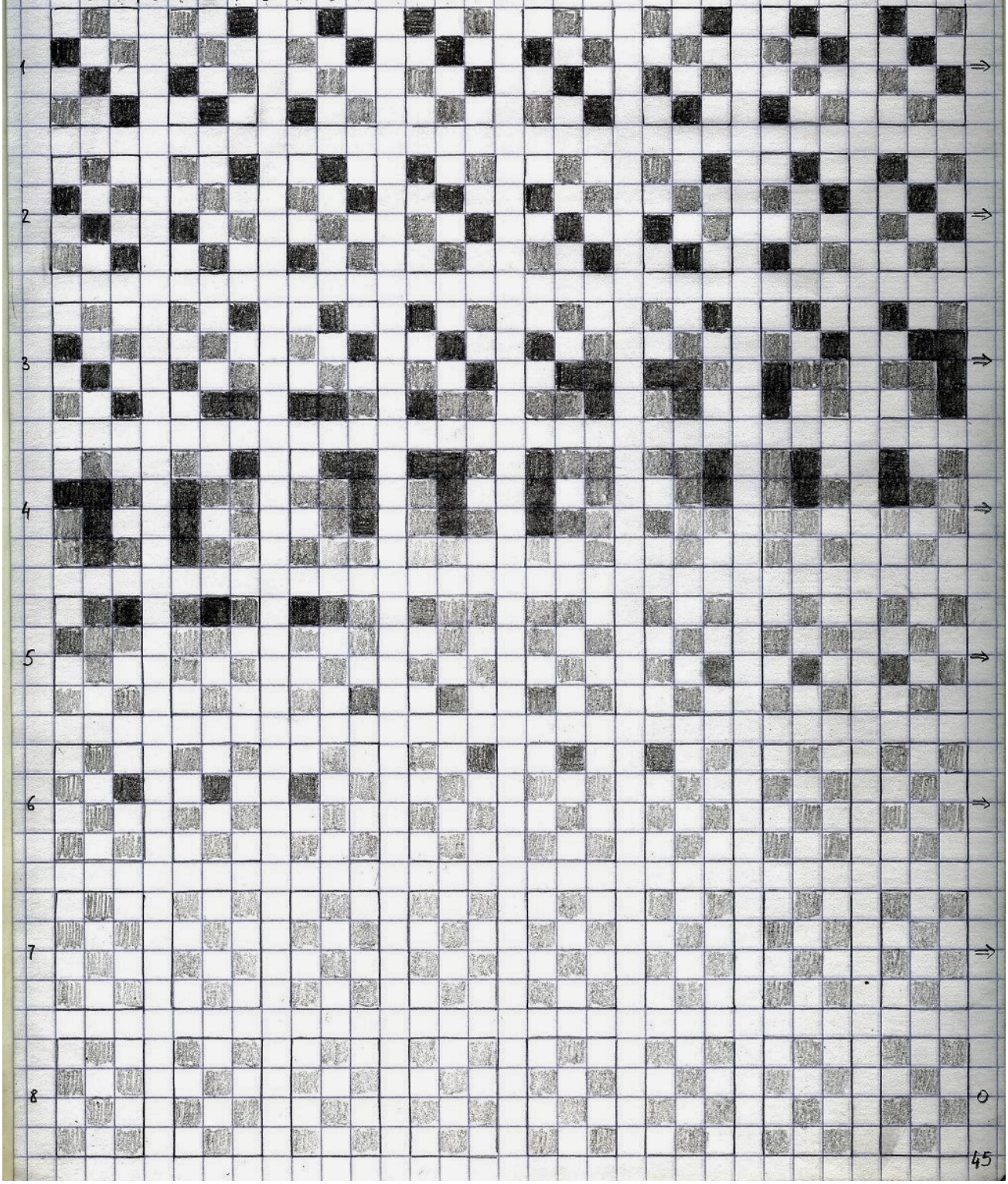


P.63251: TATACACACACACACACACACACACACAC
ACACACACACACATATATATATATATATATATA
TATATATATATATATATATA

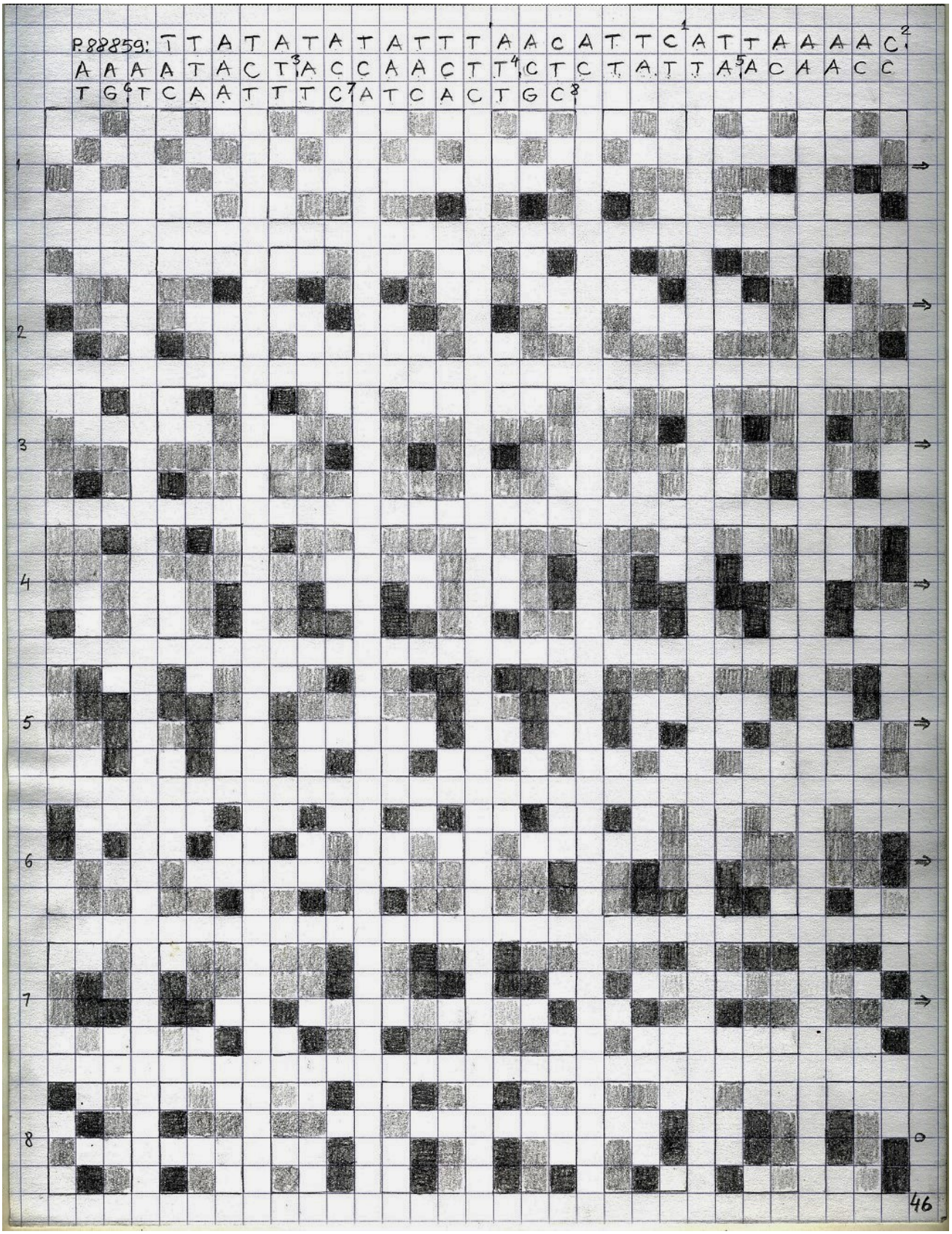


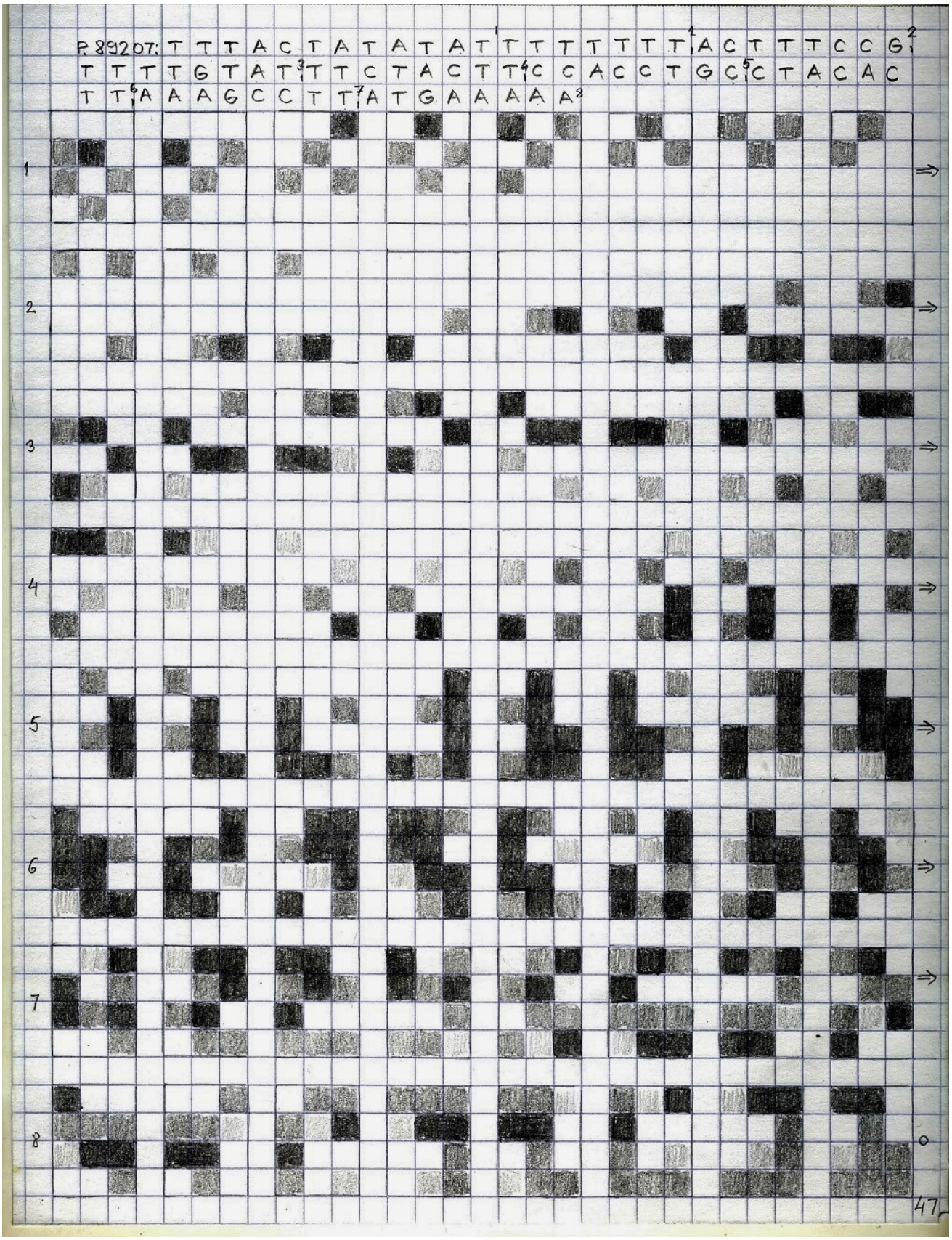

P.66984: TCTTGAATTTGTCTTCAACCTTAATTC
TTTTTTTTTATTTTTTACAATTTTTCTTTGA
AACTTTCAGTAGTTTGTA
1
2
3
4
5
6
7
8
44

P.79398: TATCTATCTATCTATCTATCTATCTATCTAT
CCAACTACAGGTGTGTGTATGTGTGTGTGT
GTGTGTGTGTGTGTGTGTGTGT

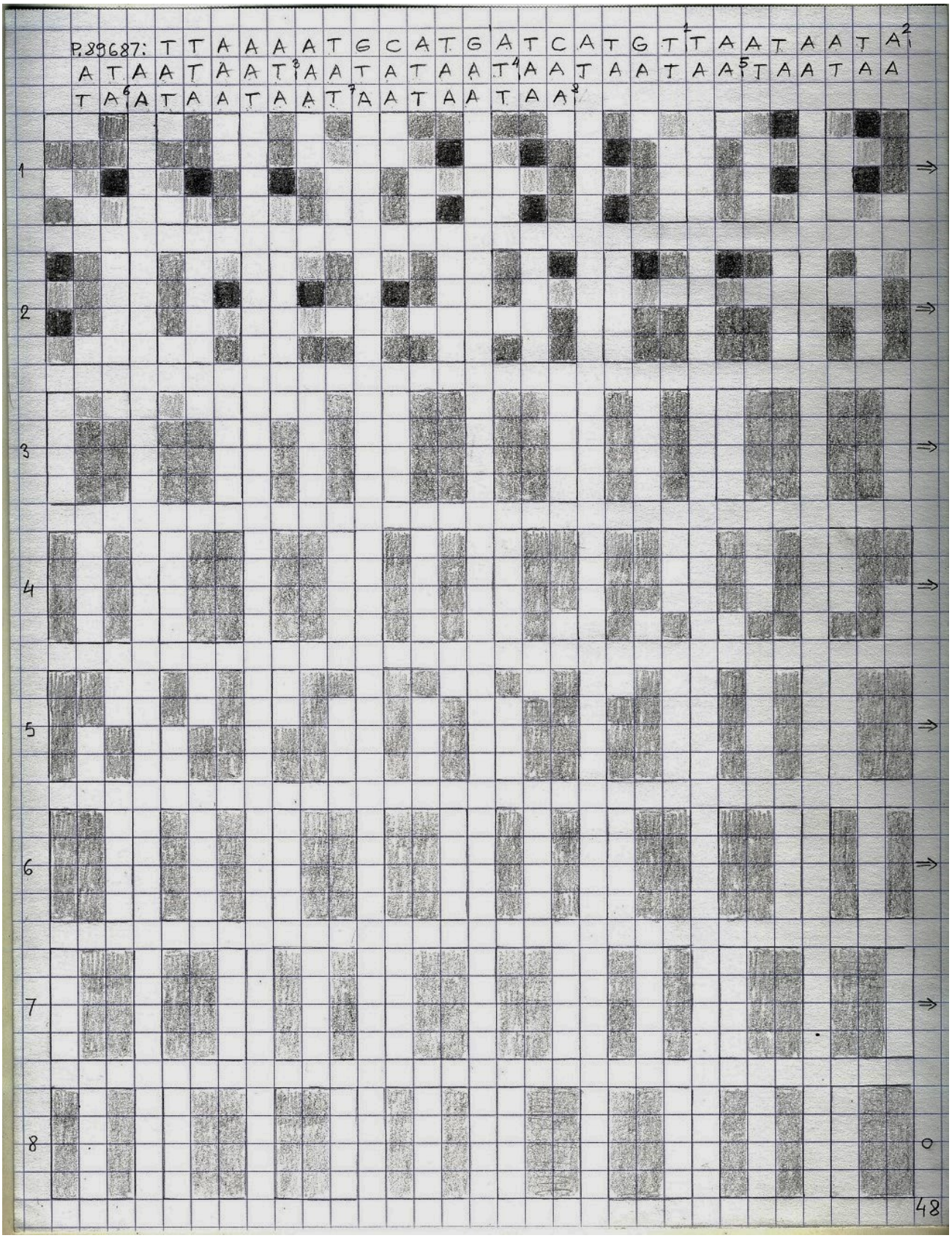


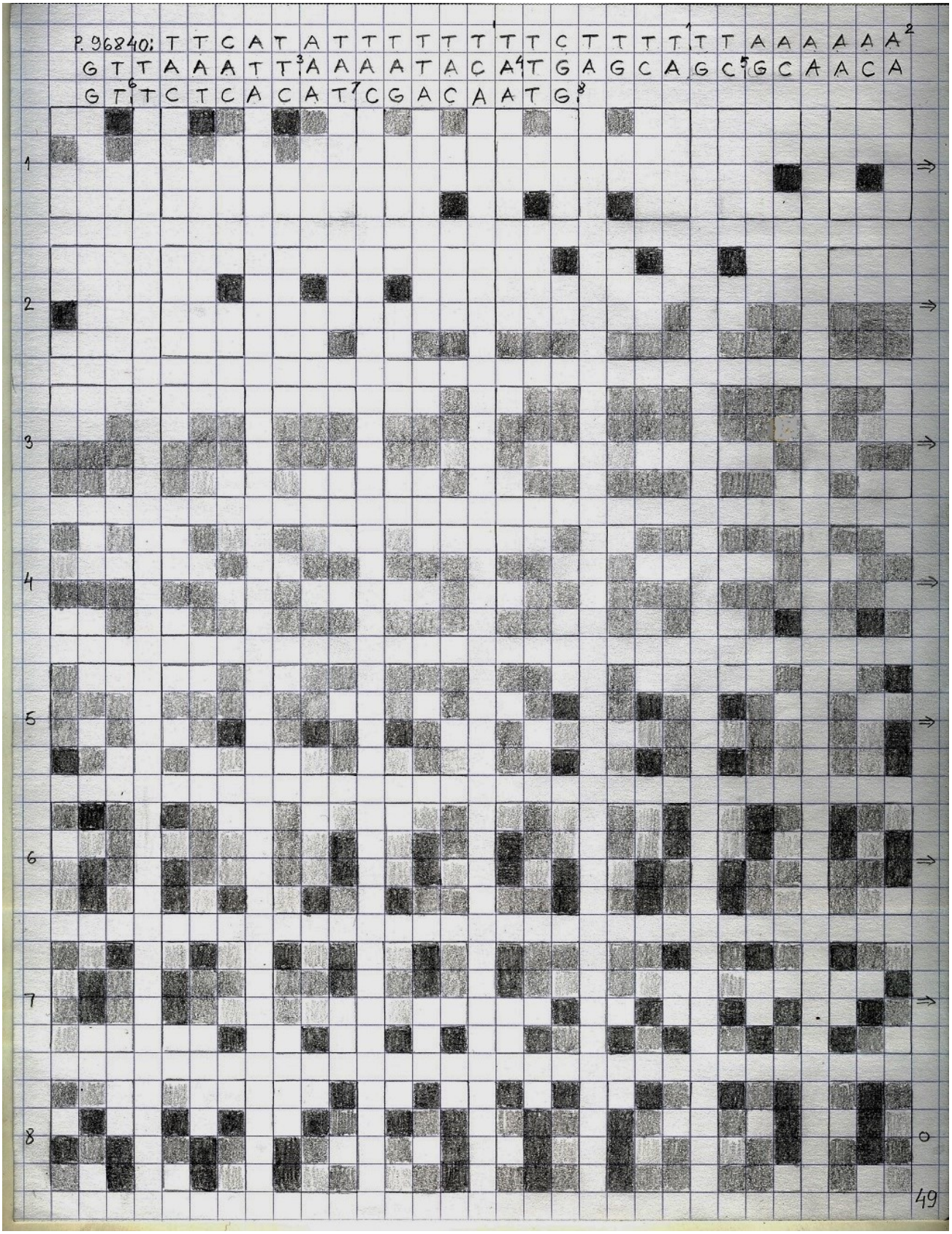
P.88859: TTATATATATTTAACATTC[1]ATTAAAAAC[2]
AAAATACT[3]ACCAACTT[4]CTCTATTA[5]ACAACC
TG[6]TCAATTTC[7]ATCACTGC[8]

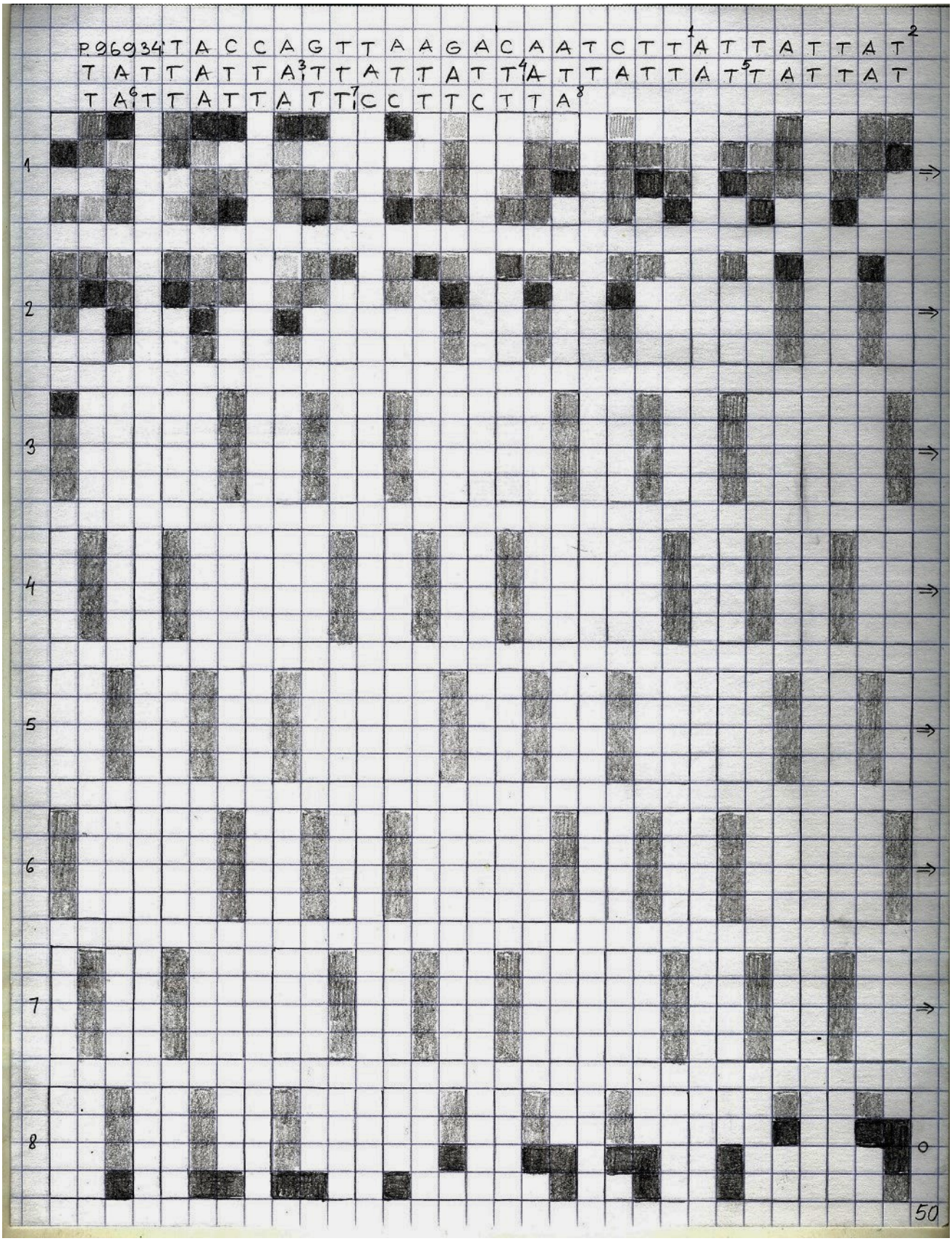


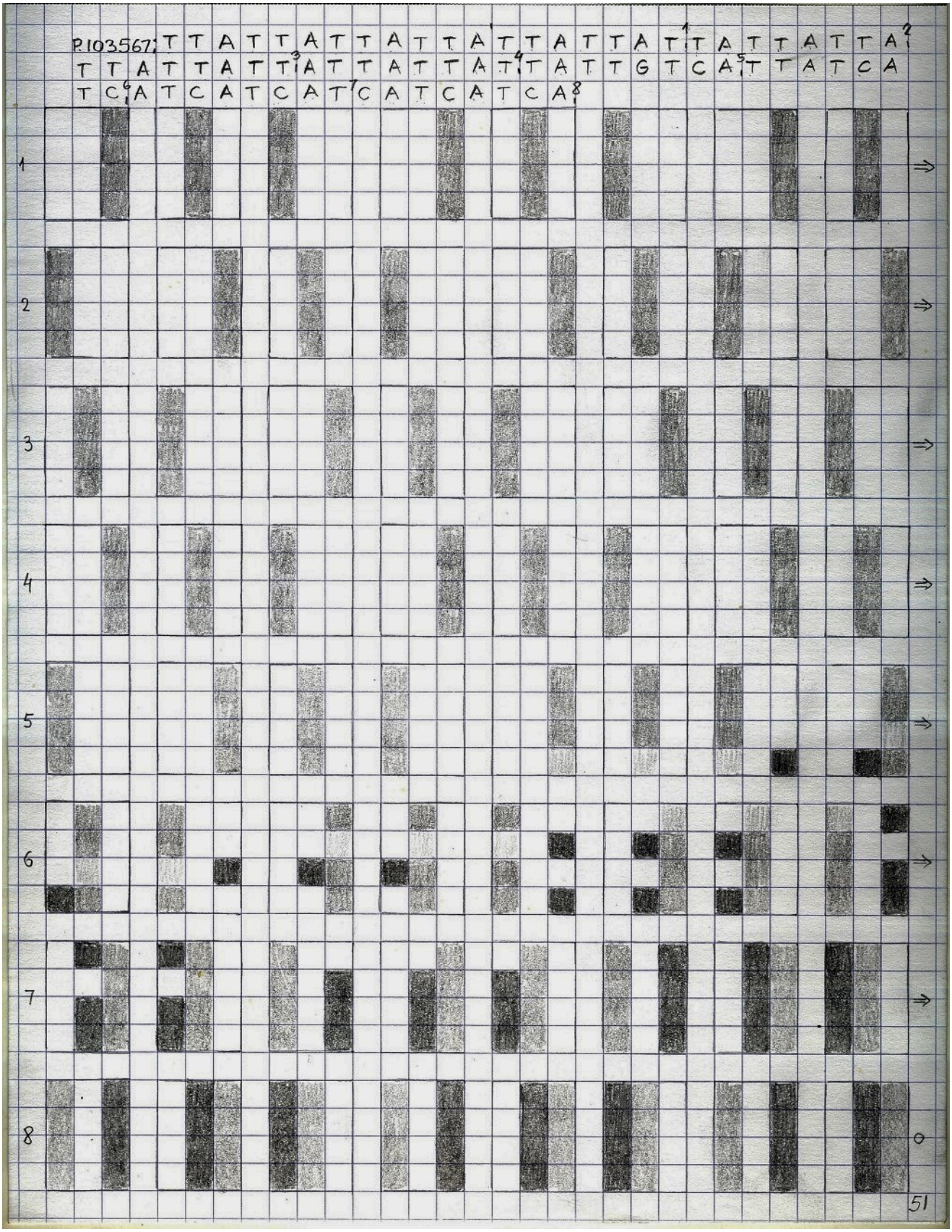

P.89207: TTTACTATATATTTTTTTTACTTTCCG
TTTTGTATTTCTACTTCCACCTGCCTACAC
TTAAAAGCCTTATGAAAAA
1
2
3
4
5
6
7
8

P.89687: TTAAAATGCATGATCATGTTAATAATA
ATAATAATAATAATAATAATAATAATAATAA
TAATAATAATAATAATAATAATAA

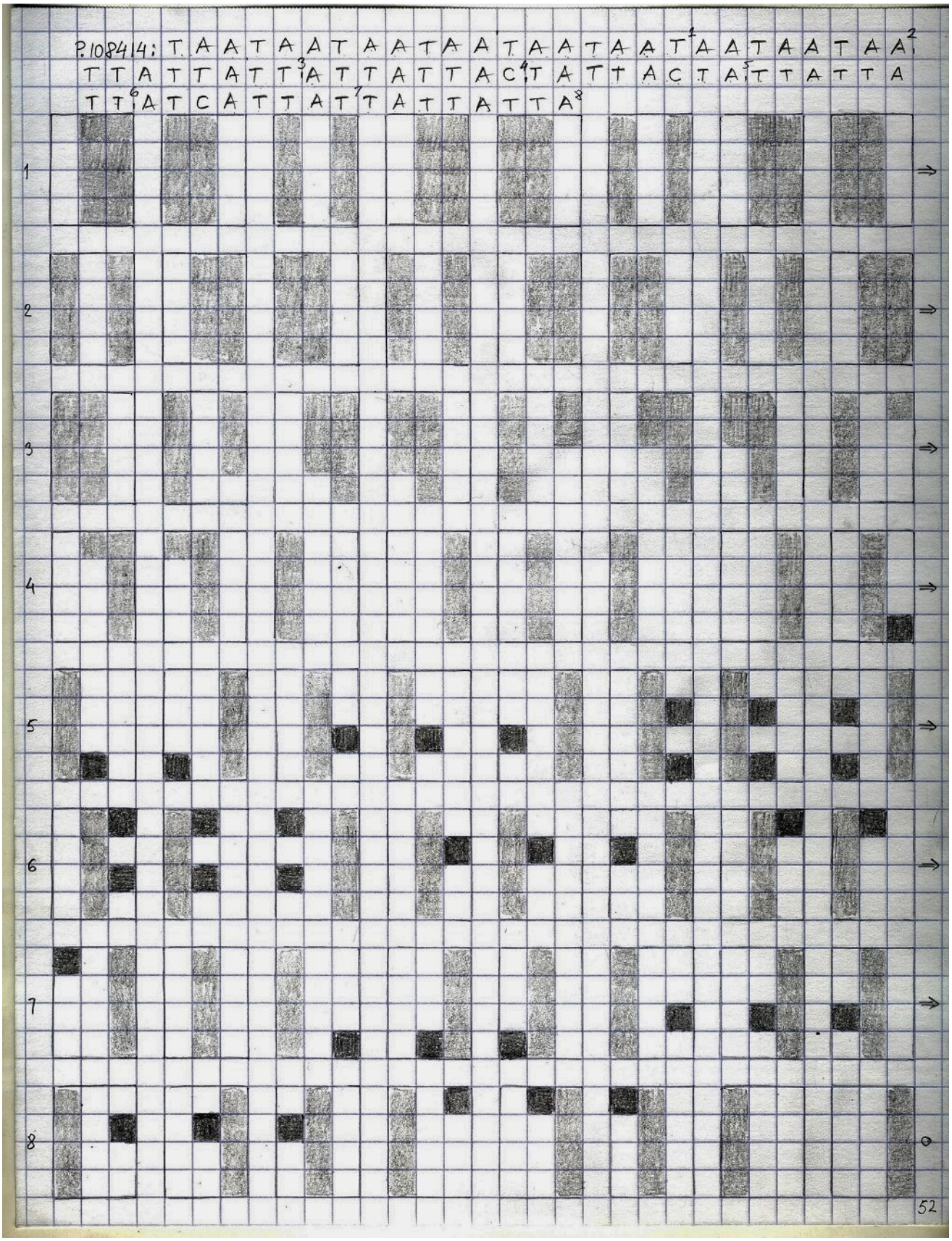


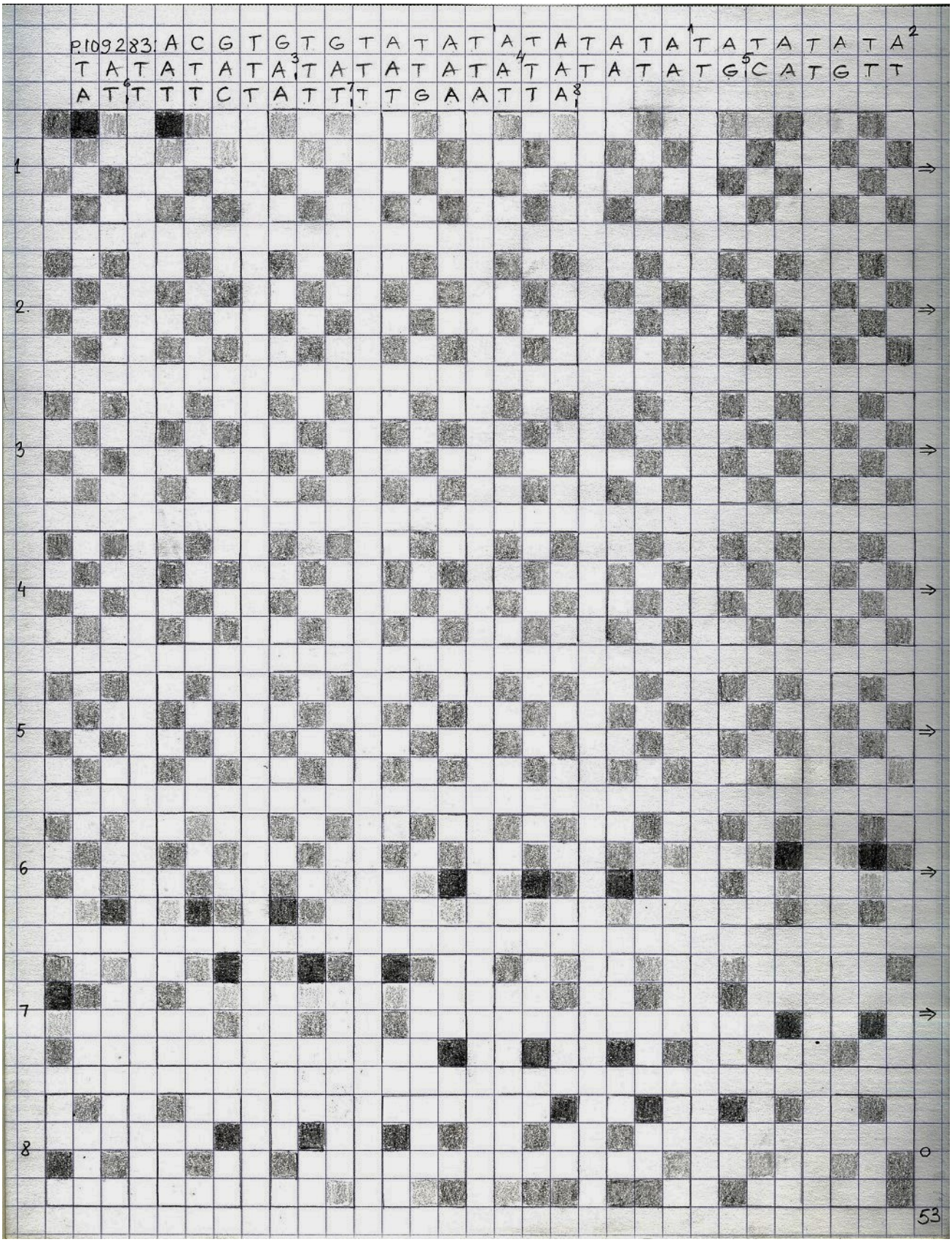

P. 96840: TTCATATTTTTTTTTTCTTTTTTAAAAAAA
GTTAAATTAAAATACATGAGCAGCGCAACA
GTTCTCACATCGACAATG
1
2
3
4
5
6
7
8

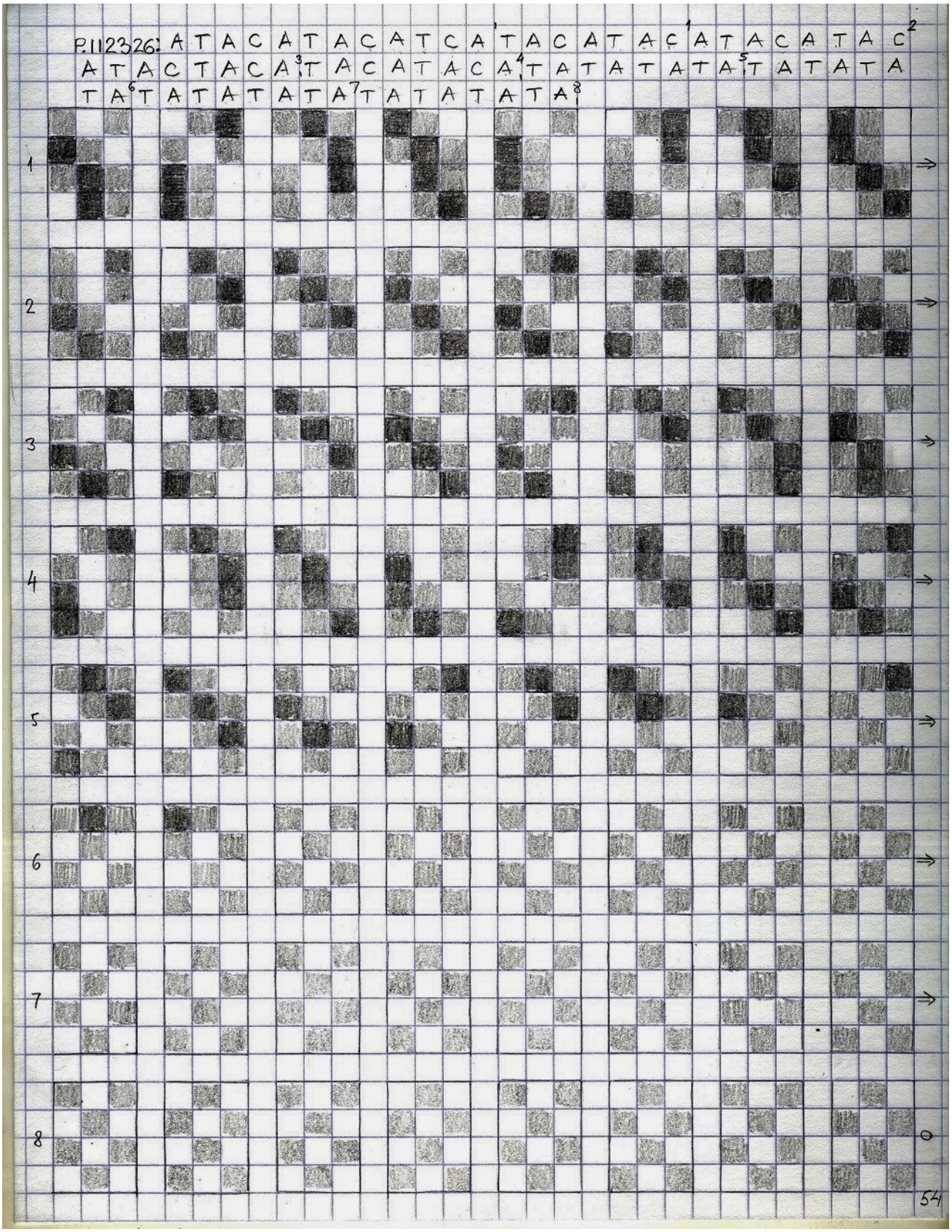

P96934 TACCAGTTAAGACAATCTTATTATTATTAT
TATTATTATTATTATTATTATTATTATTATTAT
TATTATTATTCCTTCTTA
1
2
3
4
5
6
7
8
50

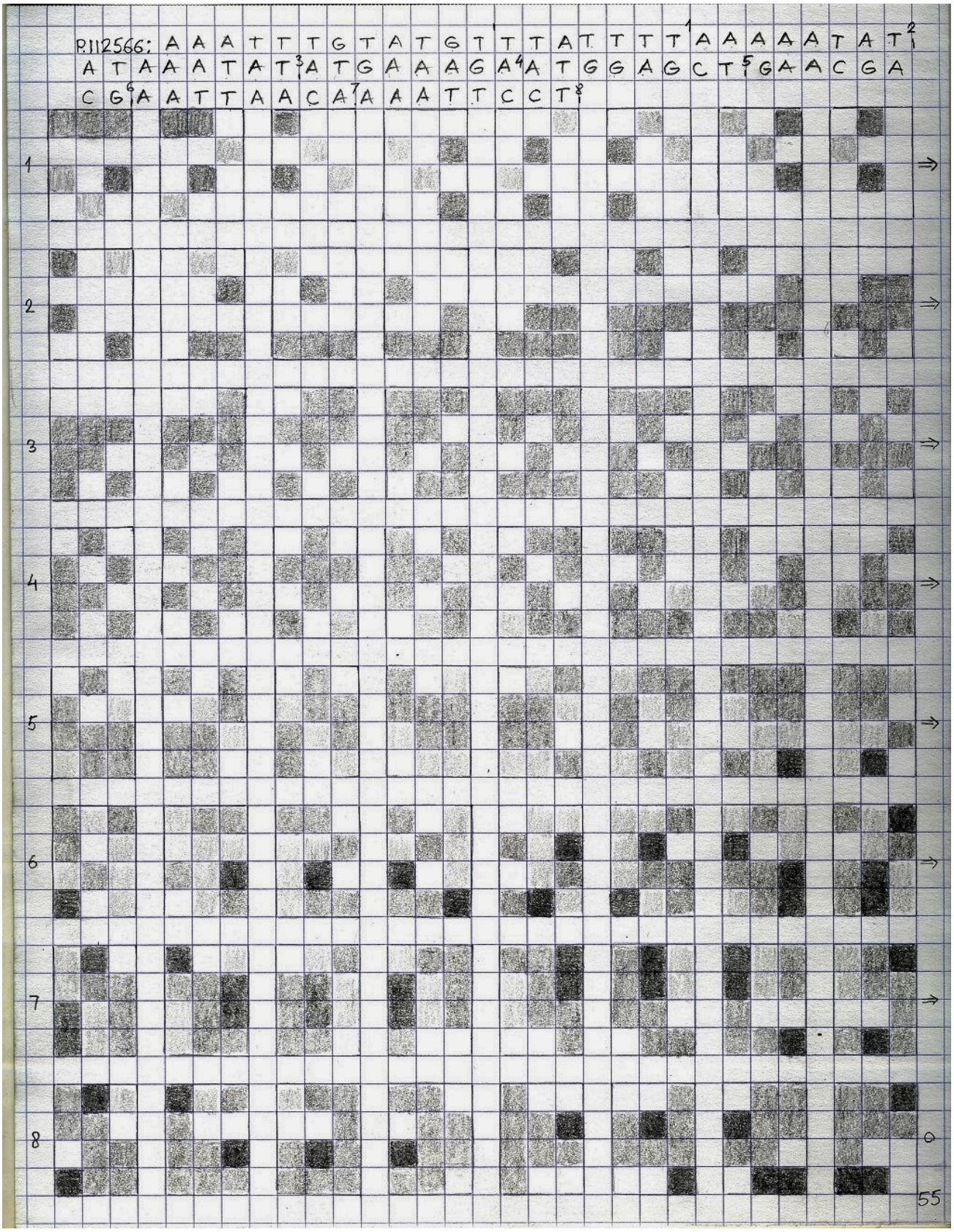

P.103567: T T A T T A T T A T T A T T A T T A T T A T T A T T A T T A T T A
T T A T T A T T A T T A T T A T T A T T A T T G T C A T T A T C A
T C A T C A T C A T C A T C A T C A T C A T C A
1
2
3
4
5
6
7
8
51

P.108414: TAATAATAATAATAATAATAATAATAATAATAA
TTATTATTATTACTATTACTATTATTA
TTATCATTATTATTATTA

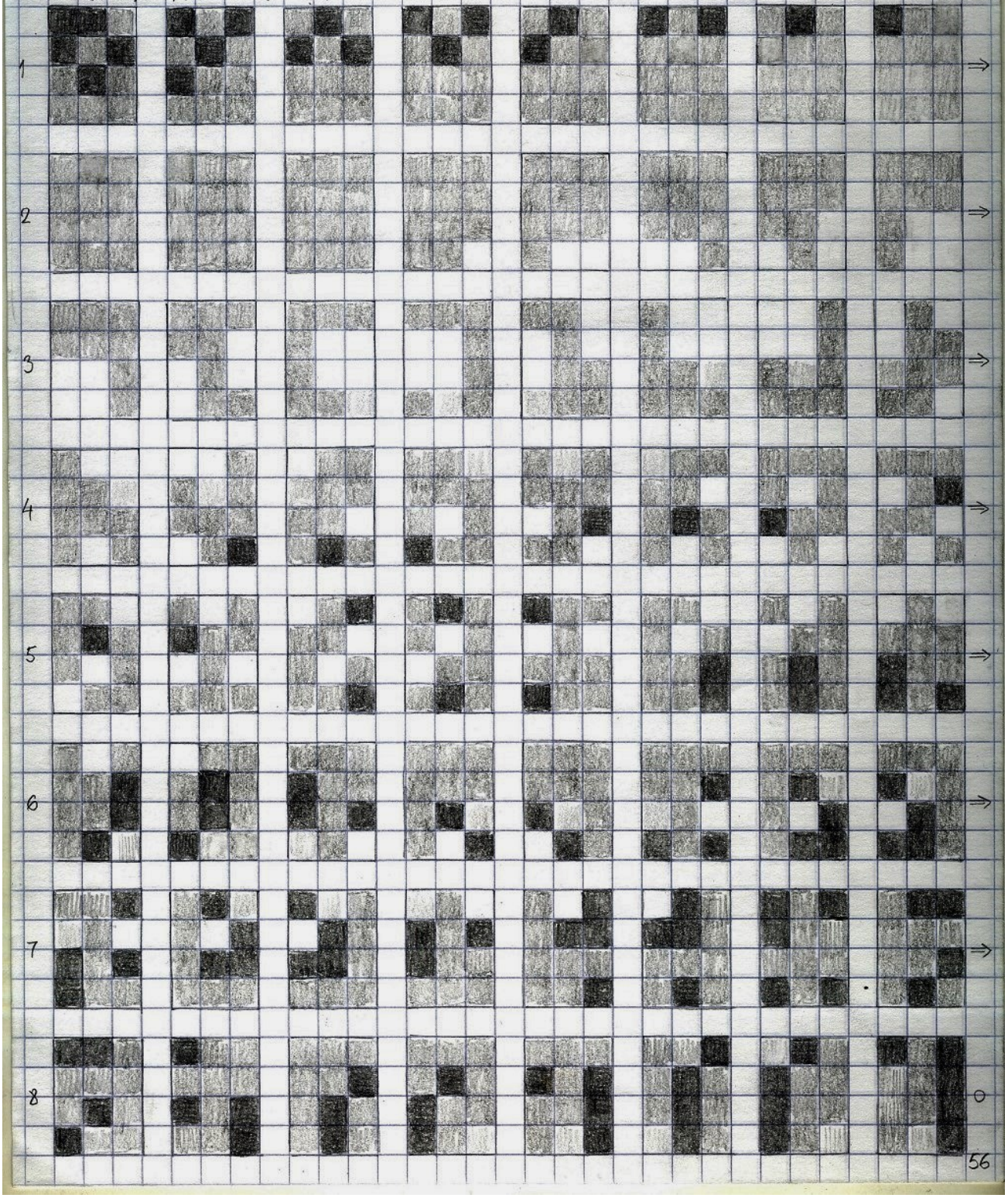


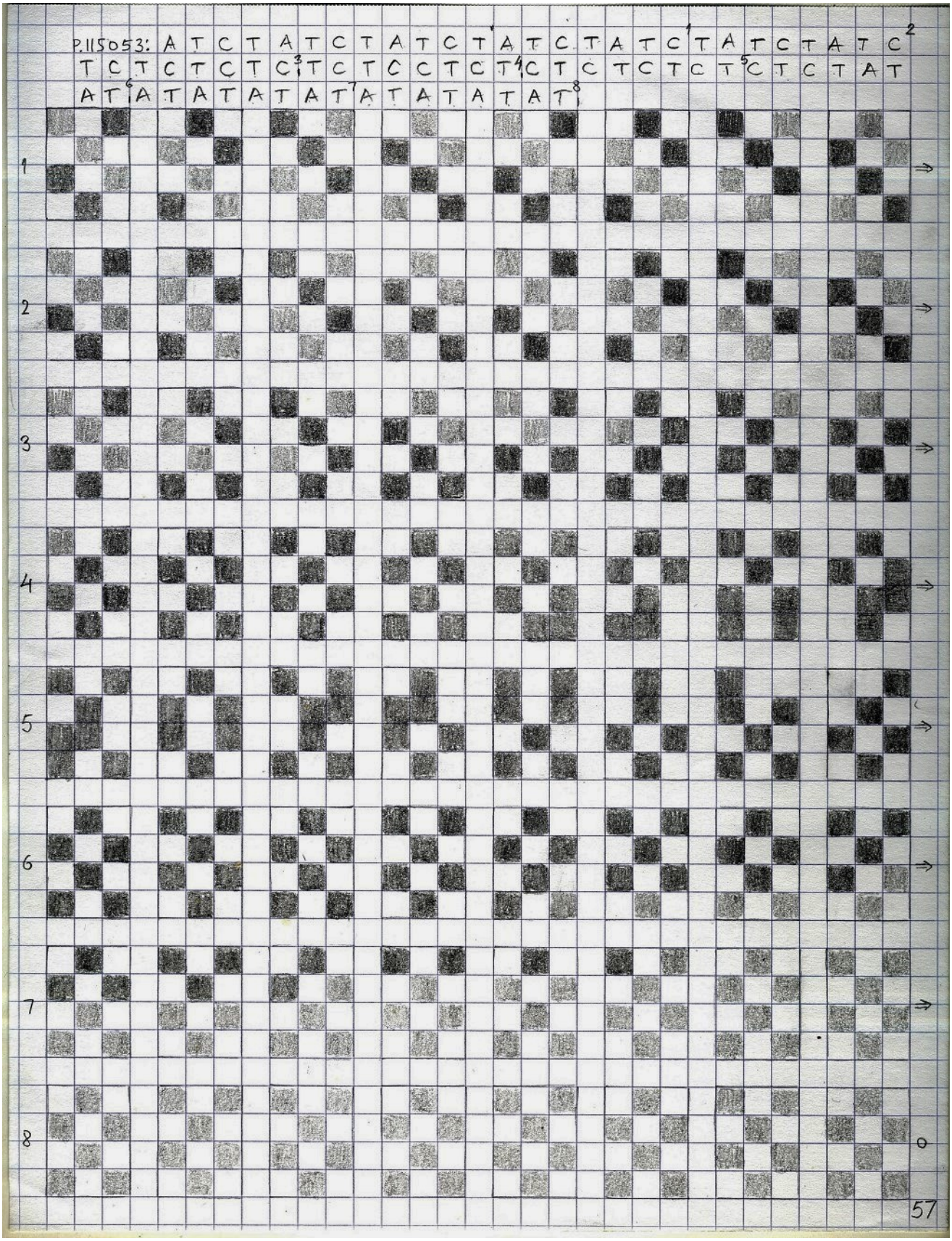

P.109283: ACGTGTGTATATATATATATATATA
TATATATATATATATATATATGCATGTT
ATTTTCTATTTTGAATTA
1
2
3
4
5
6
7
8
53

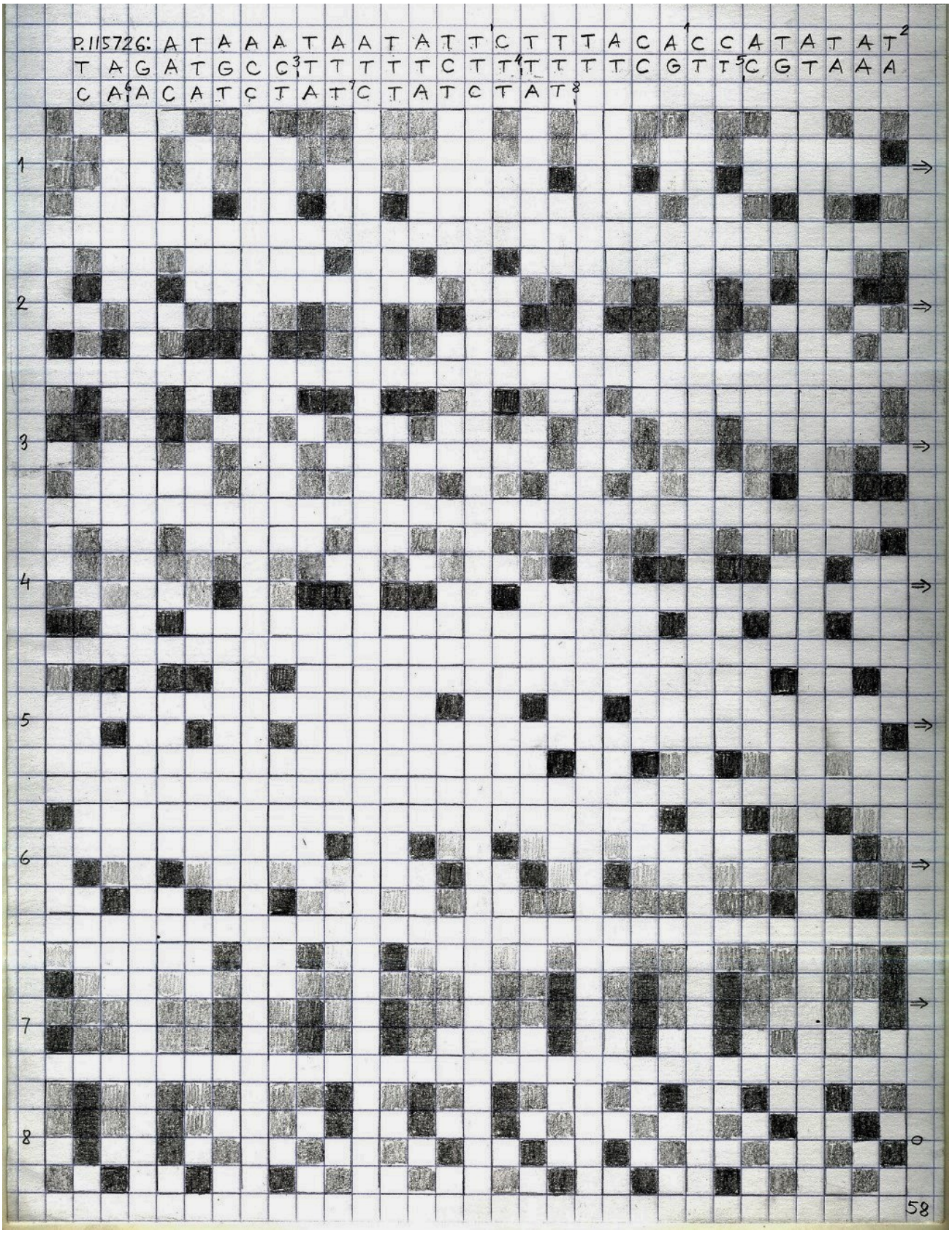

P.112326: ATACATACATCATACATACATACATAC
ATACTACATACATACATATATATATATATA
TATATATATATATATATATA
1
2
3
4
5
6
7
8
54

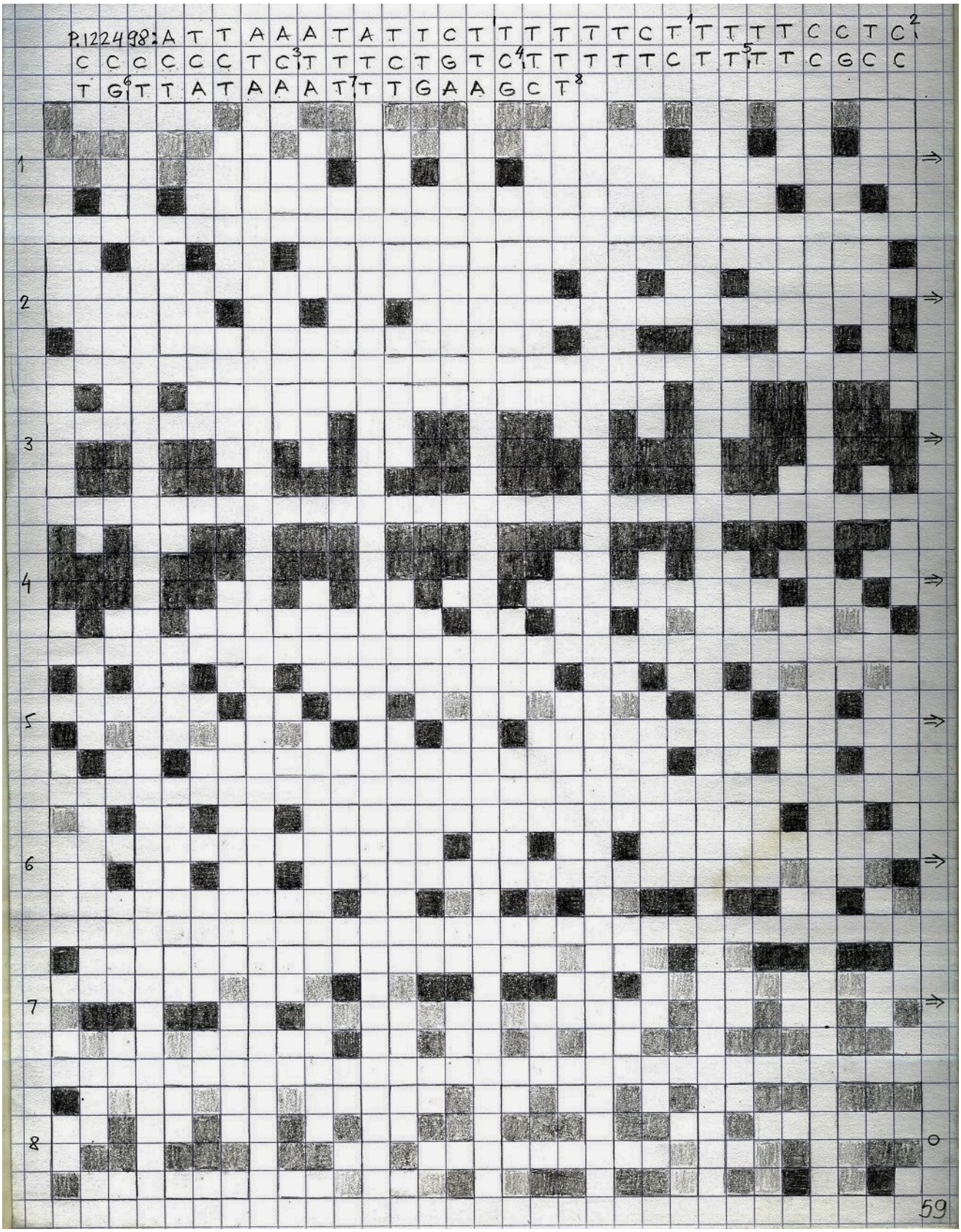

P.112566: AAATTTGTATGTTTATTTTAAAAATAT
ATAAATAT ATGAAAGA ATGGAGCT GAACGA
CG AATTAACA AAATTCCT
1
2
3
4
5
6
7
8
55

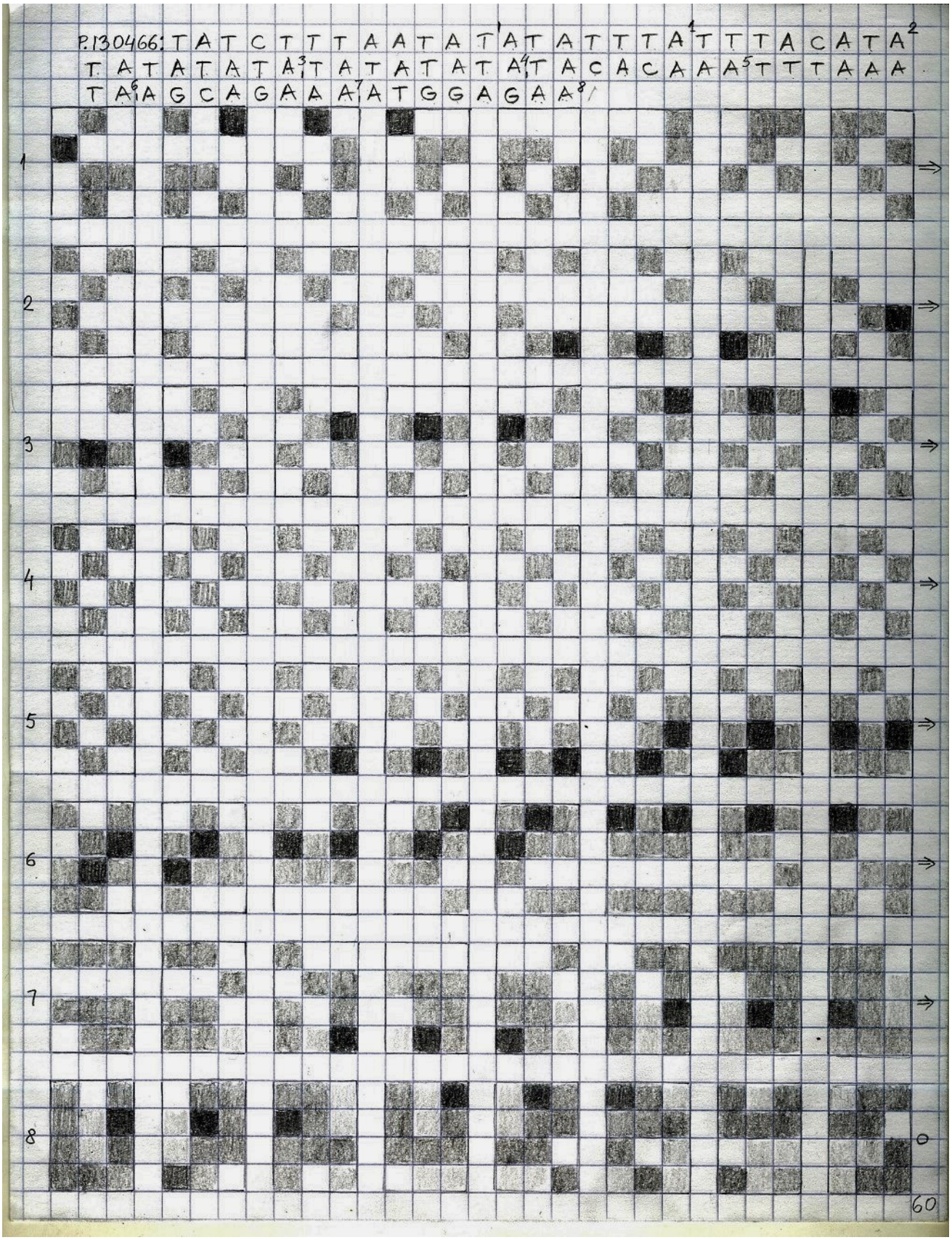
P.114713: C C A C A C A C A A A A A A A A A A A A A A A T T A T T
A A G A A A A T A C A A T A T A A A C A A C A C G A T C A C
C A A A A A C A C A A C A G C A G C




P.115053: ATCTATCTATCTATCTATCTATCTATC
TCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTAT
ATATATATATATATATATATATATATATATATA
1
2
3
4
5
6
7
8


P.115726: ATAAATAATATTCTTTACACCATATAT
TAGATGCCTTTTTCTTTTTTCGTTCGTAAA
CAACATCTATCTATCTAT
1
2
3
4
5
6
7
8
58


P.122498: A T T A A A T A T T C T T T T T C T T T T C C T C
C C C C C C T C T T T C T G T C T T T T C T T T T C G C C
T G T T T A T A A A T T T T G A A G C T
1
2
3
4
5
6
7
8
59


P.130466: TATCTTTAATATATATTTATTTACATA
TATATATATATATATATATACACAAATTTAAA
TAAGCAGAAAATGGAGAA
1
2
3
4
5
6
7
8
60